ESCHERICHIA COLI PROTEOMICS AND BIOINFORMATICS

A Dissertation

by

LILI NIU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

May 2007

Major Subject: Biochemistry

ESCHERICHIA COLI PROTEOMICS AND BIOINFORMATICS

A Dissertation

by

LILI NIU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	James C. Hu
Committee Members,	Michael Polymenis
	Ryland Young
	David H. Russell
Head of Department,	Gregory D. Reinhart

May 2007

Major Subject: Biochemistry

ABSTRACT

Escherichia coli Proteomics and Bioinformatics.

(May 2007)

Lili Niu, B.Sc., Beijing University

Chair of Advisory Committee: Dr. James C. Hu

A lot of things happen to proteins when *Escherichia coli* cells enter stationary phase, such as protein amount, post-translational modifications, conformation changes, and component of protein complex. Proteomics, which study the whole component of proteins, can be used to study the products of the genome and the physiology of *Escherichia coli* cells at different conditions. By comparing proteome from different growth phases, such as exponential and stationary phase, a lot of proteins with changes can be identified at the same time, which provides a pilot for further studies of mechanism.

Current global proteomic studies have identified about 27% of the annotated proteins of *E. coli*, most of which are predicted to be abundance proteins. Subproteomics, the study of specific subsets of the proteome, can be used to study specific functional classes of proteins and low abundance proteins. In this dissertation, using non-denatured anion exchange column with 2D SDS-PAGE and tandem mass spectrometry, difference of *E. coli* cells between exponential and stationary phase were studied for whole soluble proteome. Also, using heparin column and mass spectrometry with tandem mass spectrometry, heparin-binding proteins were identified and analyzed for exponential

growth and stationary phases. To manage and display the data generated by proteomics, a web-based database has been constructed for experiments in *E. coli* proteomics (EEP), which includes NonDeLC, Heparome, AIX/2D PAGE and other proteomic studies.

DEDICATION

To my father Jingang Niu

ACKNOWLEDGEMENTS

There are many people I would like to thank for my graduate study.

First, I would like to thank to my supervisor, Dr. James Hu, for supporting me. With his patience, his inspiration and his enthusiasm, he led me to learn useful knowledge, to solve problems, and most importantly to think. He encouraged me to learn bioinformatics, which is a powerful tool that opened my mind to a new world of biological study. Without his help, I would never finish.

I would like to thank members of my committee, Dr. Michael Polymenis, Dr. Ryland Young, and Dr. David Russell. They were always supportive, helpful and gave me valuable suggestions for my research. I especially thank Dr. Ryland Young for patiently answering my questions everyday when I started my project and offering me an RA position to finally support my study.

Also, I would like to thank Dr. Deborah Siegele, who is so nice and knowledgeable. She always took her valuable time to give me suggestions, help me solve problems, and encourage me. I appreciate her spending tremendous time to help me during my dissertation writing.

For our collaboration, I would like to thank Dr. Sam Perkins for running mass spectrometry and Dr. Larry Dangott for running 2D PAGE for us.

I would like to thank current and previous lab members, Ritchie Jabil, Jon Minor, Gregory Munoz, Adrienne Zweifel, Gwen Knapp, Dr. Leonardo Mariño-Ramírez and Dr. Matthew Champion for teaching me, giving me advice and spending time with me not

only in the lab but also in my life.

Finally, I would like to thank my friends and family. Thanks to my Mum and Dad for giving me life and raising me. Thanks to my husband, Gang Wang, for supporting me, understanding me and accompanying me.

TABLE OF CONTENTS

Page

ABSTRACT…….. iii

DEDICATION……..... v

ACKNOWLEDGEMENTS……..……..... vi

TABLE OF CONTENTS…… ……..…….……..... viii

LIST OF FIGURES…….……..……..... x

LIST OF TABLES…….……..……..... xiv

CHAPTER

I INTRODUCTION……..…….…… 1

Overview……..…….……..…. 1
Stationary phase of *E. coli*…..…….……..…... 1
Importance of proteomic and subproteomic studies…..……... 4
Methods to study proteomics and subproteomics……..... 6
Current status of proteomics studies in *E. coli*…….....…. 9
Methods to manage and display proteomic data……..…..…... 11
Studies of this dissertation and content of each chapter..... 12

II HEPAROME……..... 14

Background……..……...…. 14
Results……..……..…..…... 16
Discussion……..……..…..…. 40
Materials and methods……..……..…..... 50

CHAPTER Page

III PROTEOMIC STUDIES BY AIX COLUMN AND 2D PAGE…… 56

Background……..……..……..…. 56
Results…….…….. 57
Discussion……..…….…….... 79
Materials and methods…..……..……..…….. 86

IV EEP…….……..... 90

Background……..…….....……..…. 90
Results…….…….…… 91
Discussion…….. 99

V BIOINFORMATICS……..….. 100

Background…….…. 100
Results…….…….…. 101
Discussion……..... 134

VI SUMMARY AND CONCLUSIONS…….…… 136

Summary……..……... 136
Discussion and future prospects…..……..…... 137

REFERENCES……..……..…….…... 139

APPENDIX A…….…….... 166

VITA……..…….……... 200

LIST OF FIGURES

FIGURE Page

2.1 The structure of heparin, a sulfated poly-glycosaminoglycan..…..... 15

2.2 Flowchart of the strategy to identify heparin-binding proteins of *E. coli* cells……..……..……..….. 17

2.3 UV_{280} traces from the AKTA HPLC heparin column for samples from exponential growth (A) and stationary phase (B) of *E. coli* cells…….... 18

2.4 SDS-PAGE separation for heparin columns fractions from exponential growth (A) and stationary phase (B) lysate of *E. coli* cells.……..… 19

2.5 Distribution of protein identification on heparin columns…….…..... 21

2.6 Comparison of heparin-binding proteins between exponential growth and stationary phase……..……..……..….. 25

2.7 Comparisons of the distributions of molecular weight (A) and pI (B) between heparin-binding proteins and predicted proteins of the whole genome…….…..…….…..... 26

2.8 Functional classification by COG for heparin-binding proteins, Genome, Swiss 2D-PAGE, Corbin's, Project CyberCell, NonDeLC (M9) and NonDeLC (MOPS)……..…..….. 29

2.9 COG functional classification of heparin-binding proteins found only in exponential or stationary phase…….…… 31

2.10 Identification of FtnA in heparin-column elution fractions by Western blotting with anti-FtnA antibody.……..……..……...…...……. 35

2.11 Distribution of difference between Obs. MW and theoretical MW for heparin-binding proteins identified by in-gel digestion combined with MS or/and MS/MS…….…….…. 36

FIGURE Page

2.12 pI distribution of heparin-binding proteins in elution fractions of the exponential growth sample... 38

2.13 pI distribution of heparin-binding proteins in elution fractions of the stationary phase sample... 39

2.14 Elution profiles of different amount protein loaded on heparin columns... 43

2.15 A) Elution positions of RelA on heparin columns. B) Distribution of ribosomal proteins on heparin columns... 45

2.16 A) Venn diagram of summary for Heparome and other proteomic studies. B) Distribution of the CAI range of proteins identified by current proteomic studies. ... 48

3.1 Flowchart of the strategy used to identify the soluble proteome of *E. coli* by LC and 2D PAGE... 59

3.2 Analysis of 2D gels using Z3 2D gel analysis system... 60

3.3 2D PAGE of fractions from AIX HPLC for both exponential growth and stationary phase... 62

3.4 Strategy to analyze a series of 2D PAGE... 63

3.5 Example of elution profiles for spots visualize on a series of 2D PAGE... 65

3.6 Elution profiles of proteins with pI from 7 to 8 and with MW from 20 to 200 kD... 66

3.7 Information of the protein spot identified as DnaK on 2D PAGE... 73

3.8 Information of protein spots identified as AceA on 2D PAGE... 74

3.9 Comparison of elution profiles between 2D PAGE and 2DLC... 77

3.10 The virtual 2D PAGE (A), the real 2D PAGE (B) and the redrawn real 2D PAGE (C) for 6th elution fraction from AIX column for exponential growth sample of *E. coli*... 78

FIGURE Page

4.1A Structure of the MySQL database of heparin-binding proteins (Heparome)…….. 92

4.1B Structure of the MySQL database of NonDeLC and other proteomic studies…….. 93

4.2 Connections between the webpage interface and the MySQL databases……..... 94

4.3 The website content for NonDeLC……..... 96

4.4 Comparison of elution patterns between DapA and DapB on NonDeLC…… 97

4.5 Information of DnaK identified by NonDeLC…….. 98

5.1 The genome view of *E. coli* K-12 MG1655 in Gbrowse…… 105

5.2 The genome view of *E. coli* K-12 W3110 in Gbrowse……..... 107

5.3 The Gbrowse details page for *dnaK* in version of Genobase 2005 of *E. coli* K-12 W3110…….. 109

5.4 Genes with *groP* as synonyms in *E. coli* K-12 MG1655…… 110

5.5 The form to submit ORF searching in GetORF…….. 113

5.6 Results of ORF searching for the complete genome of *Burkholderia cenocepacia* phage Bcep1 by GetORF …….. 114

5.7 The form to submit lipoproteins searching in LipoRy……. 117

5.8 Results of lipoprotein searching for the complete genome of bacteriophage lambda……... 118

5.9 Results of *Rz/Rz1* pair searching from enterobacteria phase lambda using GetRzRz1…… 121

5.10 The result of SAR endolysin searching from genome of enterobacteria phage P1……... 123

5.11 The form to display primary structure of protein sequence……... 125

FIGURE Page

5.12 Display of primary structure with charges for serine sensor receptor of *E. coli* K-12……..... 126

5.13 Calculation of amino acid content of Rz1 proteins of lambdoid phages versus the total proteins of *E. coli* ……..……...……. 127

5.14 The form for "GroupProtein"……..…….……. 129

5.15 Results of BLASTing and grouping APSE-1 Rz1, ShigellaV Rz1, and Mu Rz 1 using "GroupProtein". …….……... 130

5.16 Phage genomes with candidates of lysis gene(s)…….…... 132

5.17 Details of every candidate of phage lysis genes from annotated phage genomes……..… 133

LIST OF TABLES

TABLE Page

1.1 *E. coli* web-based databases 13

2.1A Heparin-binding proteins identified from exponential growth and stationary phase *E. coli* cells..... 22

2.1B Heparin-binding proteins found in exponential growth of *E. coli* cells.. 23

2.1C Heparin-binding proteins found in stationary phase of *E. coli* cells..... 24

2.2 Heparin-binding proteins with abundance changes between exponential growth and stationary phase..... 32

2.3 Proteins with elution shift on the heparin column between exponential growth and stationary phase..... 34

2.4 Genes of transcription regulators in *E. coli* K-12..... 49

3.1 Number of spots visualized on every 2D PAGE..... 61

3.2A Cut spots with protein identified..... 67

3.2B Cut spots without protein identified..... 68

3.3 Proteins identified from 2D PAGE..... 69

3.4 The summary of proteins identified from 2D PAGE..... 75

3.5 Proteins more abundant in stationary phase identified by 2D PAGE.... 82

5.1 Annotation tracks for *E. coli* Gbrowse..... 106

A.1 Heparin-binding proteins identified in every fraction from heparin column..... 166

CHAPTER I

INTRODUCTION

Overview

In chapter I, the stationary phase of *E. coli*, importance of proteomic and subproteomic studies, current status of proteomics studies, methods to study proteomics and methods to manage and display proteomic data will be discussed. Then the objectives of this dissertation and the content of each chapter will be described.

Stationary phase of *E. coli*

E. coli is one of the most commonly used model organisms for molecular biology. In nature, *E. coli* can be found in the intestinal tract of warm-blooded animals, groundwater, and spring. *E. coli* can grow in a variety of conditions, such as low nutrient (Death and Ferenci, 1994), abundant nutrient, high pH (Avetisyan *et al.*, 1991), low pH (Hsin-Yi and Chou, 2001; Leyer *et al.*, 1995), aerobic (Iuchi and Weiner, 1996), anaerobic (Partridge *et al.*, 2006; Weber *et al.*, 2006), and wide temperature range (Cullum *et al.*, 2001; Riehle *et al.*, 2005), many of which are unlike the growth conditions commonly used in the laboratory.

This dissertation follows the style of *Molecular Microbiology*.

In the laboratory, the first growth phase of cells is lag phase when the growth rate is slow because cells are trying to adapt to the new growth condition. Once cells adapt to the new condition, cells enter exponential phase when they multiply exponentially. Stationary phase is the growth phase when the growth rate slows due to depletion of nutrients and accumulation of toxic products (Ramirez Santos *et al.*, 2005). When *E. coli* cells enter stationary phase in the laboratory, many conditions are similar to the natural growth conditions, which makes stationary phase a good example to study how *E. coli* cells adapt to changes and stresses in nature.

During the transition from exponential growth to stationary phase, *E. coli* cells have many changes in both morphology and physiology and become more resistant to stresses such as starvation, oxidative stress, osmotic stress, and low pH (Ishihama, 1997). In stationary phase, cells become more spherical and smaller, which is due to induction of the *bolA* gene (Loewen and Hengge-Aronis, 1994). BolA is very likely affecting cell morphology by controlling transcription of *dacA*, *dacC*, and *ampC* (Hengge-Aronis, 2002). The cell membrane becomes more rigid (Ramirez Santos *et al.*, 2005). Stationary phase cells do not divide their genetic material until all cells reach one single chromosome.

Changes also occur in the genetic material. The nucleoid is more condensed (Ishihama, 1997). Many defects arise on DNA such as fragmentation of chromosome, deletions and point mutation. mRNAs for stationary phase-specific genes, such as *rpoS*, *bo1A*, and *dps*, are polyadenylated (Cao and Sarkar, 1997). Synthesis of ribosomal RNA

is shut off, which is not regulated by Fis, H-NS and RpoS (Aviv *et al.*, 1996).

In stationary phase, changes also happen to proteins. While expression of many proteins is decreased (Ishihama, 1997), others are up-regulated (Groat *et al.*, 1986; Ishihama, 1997). RpoS, (p) ppGpp, cyclic AMP (cAMP)-CRP and Lrp regulate expression of many proteins during stationary phase (Eisenstark *et al.*, 1996; Gentry *et al.*, 1993; Hengge-Aronis, 1993, 1996, 2002; Ishihama, 1997; Lange *et al.*, 1995; Tani *et al.*, 2002). In stationary phase, RpoS (σ^S) accumulates, bind and directs the RNA polymerase to over 50 genes to response to starvation and stresses (Hengge-Aronis, 2002). During transition from exponential to stationary phase, *E. coli* cells start to produce (p)ppGpp (Kramer *et al.*, 1981), which serves as an alarm responding low nutrient in growth media. (p)ppGpp affects different cellular processes, such as DNA replication and cell division (Schreiber *et al.*, 1995), cyclopropane fatty acid synthesis (Eichel *et al.*, 1999) and gene expression. (p)ppGpp regulate expression of different genes, such as DNA replication inhibitor gene (*cspD*) (Yamanaka and Inouye, 1997), thioredoxin gene (Lim *et al.*, 2000), ribosome modulation factor (RMF) gene (Izutsu *et al.*, 2001a), ribosome-associated protein (Izutsu *et al.*, 2001b) and glutaredoxin 2 (Grx2, encoded by grxB) (Potamitou *et al.*, 2002). In stationary phase, expression of many genes is up-regulated by cAMP, such as glycogen biosynthesis genes (Romeo and Preiss, 1989), and galactose ABC transporter gene (mal) (Death and Ferenci, 1994).

When cells transit from exponential growth to stationary phase, proteins also have changes in post-translational modifications (Desnues *et al.*, 2003; Dukan and Nystrom, 1998; Fredriksson *et al.*, 2005) or conformation. The composition of some protein

complexes might also change. Specific *E. coli* proteins are oxidative modified during stationary phase. These proteins are Hsp70 chaperone DnaK, histone-like protein H-NS, elongation factor EF-Tu and EF-G, the universal protein UspA, glutamate synthase, glutamine synthetase, aconitase, malate dehydrogenase, and pyruvate kinase (Dukan and Nystrom, 1998; Tamarit *et al.*, 1998).

Importance of proteomic and subproteomic studies

Methods that have been used to study the physiology of *E. coli* cells can be grouped into two major categories: study of a single gene or gene product or study of many genes or gene products at the same time. The later approach includes genomics (Borodovsky, 1998; Dobrindt, 2005; Martin and Rosner, 2002; Mori, 2004; Riley and Serres, 2000; Tao *et al.*, 1999), transcriptomics, proteomics (Hatzimanikatis *et al.*, 1999), metabolomics (Verhoeckx *et al.*, 2004) and fluxomics (Wiback *et al.*, 2004).

Transcriptomics have been applied to study transcription *E. coli* genes under different conditions, such as nitrogen starvation (Kabir *et al.*, 2004) , leucine-responsive regulatory protein (Lrp) regulation (Tani *et al.*, 2002), RpoS regulation (Patten *et al.*, 2004). To study expression of the final product of most genes, proteins, proteomics is a more direct way than transcriptomics since mRNA level does not always correlated with the amount and properties of proteins. Translational control, protein degradation, post-translational modifications and protein-protein interactions (Anderson and Seilhamer, 1997; Anderson and Anderson, 1998; Lottspeich, 1999) cannot be detected directly by transcriptomics.

The term “proteome” first appeared in the literature in 1995 (Kahn, 1995), although

studies of proteomes have been done since the development of two-dimensional gel electrophoresis (O'Farrell, 1975). The proteome shows different expression patterns for the same organism under different conditions or in different tissues. In *E. coli*, gene expression is tightly related to physiology. In response to various conditions, such as low pH, nutrient starvation, phage infection, and antibiotics, *E. coli* regulates gene expression in order to adapt to the changes (Hua *et al.*, 2004; Schembri *et al.*, 2003). Proteomics is a way to understand how protein expression patterns are related to the physiology of *E. coli* at the protein level.

Proteomics can be used to study many subjects, such as protein expression profiles (Baglioni *et al.*, 2003; Champion *et al.*, 2003; Ihling and Sinz, 2005; Kerner *et al.*, 2005; Kim *et al.*, 2005; Lopez-Campistrous *et al.*, 2005; Raman *et al.*, 2005; Silva *et al.*, 2006; Tonella *et al.*, 2001; Xu *et al.*, 2006), protein localization (Lopez-Campistrous *et al.*, 2005), protein topology (Daley *et al.*, 2005), protein-protein interaction (Arifuzzaman *et al.*, 2006; Butland *et al.*, 2005) and protein post-translational modification (Aebersold and Mann, 2003; Larsen and Roepstorff, 2000; Simpson and Pepperkok, 2003).

Subproteomics is a term applied to the analysis of specific sets of proteins, such as proteins with a particular post-translational modification (Demirev *et al.*, 2001), proteins at different subcellular locations (Berven *et al.*, 2006; Nouwens *et al.*, 2000; Rey *et al.*, 2005; Schaumburg *et al.*, 2004; Trost *et al.*, 2005), and even insoluble proteins (Graham *et al.*, 2006). Moreover, because global proteomics are biased towards abundant proteins (Champion *et al.*, 2003; Corbin *et al.*, 2003; Tonella *et al.*, 2001), subproteomics also has the benefit of enriching and identifying low abundance proteins (Cordwell *et al.*, 2000).

Methods to study proteomics and subproteomics

To study the proteome, efficient separation of the complex protein mixture is essential. Two traditional ways to separate protein samples for proteomic studies are one-dimensional and two-dimensional gel electrophoresis (2D PAGE) (O'Farrell, 1975). One-dimensional gel electrophoresis (Schagger and von Jagow, 1987; Wolf *et al.*, 1970) is sufficient to separate protein mixtures containing less than 100 components, while two-dimensional gel electrophoresis can separate complex protein mixtures containing thousands of components (Henzel *et al.*, 1993; Langen *et al.*, 2000; Loo *et al.*, 2001; Tonella *et al.*, 2001). 2D PAGE has advantage to study protein abundance (Rosen and Ron, 2002) and post-translational modifications (Yan *et al.*, 2002). However, 2D PAGE has some limitations. For example, detecting low abundance proteins, proteins with extreme pIs, proteins with extreme molecular weights (MW) or membrane proteins is difficult (Bunai and Yamane, 2005). Further, 2-D electrophoresis and its analysis is extremely labor intensive. Recently an advanced technique, fluorescence difference 2-D gel electrophoresis technology (2D-DIGE), has been developed to improve separation and relative quantitation of complex protein mixtures (Yan *et al.*, 2002) from proteome samples with different treatments. This method is more accurate quantitative analysis than conventional 2D PAGE.

Liquid chromatography, such as two-dimensional LC separation coupled with MS, has been applied as a new tool to separate mixtures of protein (Badock *et al.*, 2001; Champion *et al.*, 2003; Ferguson and Smith, 2003) or peptides (Ihling and Sinz, 2005; Vollmer *et al.*, 2003; Yates *et al.*, 1993). Multidimensional protein identification

technology (MudPIT) is one of the popular methods which analyze large-scale proteome by multidimensional liquid chromatography and tandem mass spectrometry (Washburn *et al.*, 2001). Using MudPIT, proteome is digested into peptides first proceeding LC separation. Protein with extreme pI, MW or low abundance can be identified by LC. LC analysis of proteomes not only can overcome some of the limitations of 2-D gels, but can also keep the proteins or protein complexes in their native state (Champion *et al.*, 2003).

After separating lysates into less complex protein mixtures, individual proteins must be identified. Mass spectrometry is a sensitive and accurate way to identify proteins. Matrix assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp, 1988; Yates, 2000) and electro-spray ionization (ESI) (Duan *et al.*, 2006; Hauer *et al.*, 1992; Le Bihan *et al.*, 2001; Le Bihan *et al.*, 2003) are two commonly used mass spectrometry methods to ionize peptide to obtain precise peptide masses by coupling with different kind of mass analyzers. The majority of mass spectrometry (MS) processes utilize electrospray ionization (ESI) as a means for introducing a LC separated peptide mixture into a mass spectrometer (Wolters *et al.*, 2001). However, there are advantages for using MALDI rather than ESI for high throughput samples, specifically the ability to do the separations offline with the analysis. By having static spots "frozen in time" (Graber *et al.*, 2004) on the MALDI plate it is possible to go back and sequence specific peptides that are of interest and perform data dependent data-mining. Additionally it is often difficult with LC-ESI to determine the optimal time to sequence peptides as they elute from the LC column. Several studies have shown the tendency of MALDI to

preferentially ionize basic residues (Cohen and Chait, 1996; Krause *et al.*, 1999) but other factors such as size, secondary structure (Wenschuh *et al.*, 1998), hydrophobicity (Cohen and Chait, 1996), and charged side chains (Baumgart *et al.*, 2004) can also influence MALDI ion yields. Higher signal intensities generally produce better ion scores; however, other factors also affect the peptide fragmentation and the resultant ion score (Zhen *et al.*, 2004).

The peptide masses can be used to do MS fingerprinting by matching their peptide masses to the theoretical peptide masses generated from a protein or DNA database (Cottrell, 1994; Pappin *et al.*, 1993; Thiede *et al.*, 2005). Furthermore, by using tandem mass spectrometry (MS/MS) (Aebersold and Goodlett, 2001; Aebersold and Mann, 2003; Ferguson and Smith, 2003; Griffin *et al.*, 2001; Larsen and Roepstorff, 2000; Wu and Yates, 2003; Yates, 2000) you can get partial peptide sequences, which provide more information to increase the number and accuracy of protein identifications.

Current status of proteomics studies in *E. coli*

Current global proteomic studies have identified about 27% of the annotated proteins of *E. coli*, most of which are predicted to be abundant proteins. In the *E. coli* Swiss-2D PAGE database (Tonella *et al.*, 2001), thousands of proteins were separated by 2D PAGE gels and about 273 proteins were identified. In our lab, non-denaturing 2D LC/MS (NonDeLC) separated and identified 310 proteins (Champion *et al.*, 2003) from exponential phase of *E. coli* cells grown in M9 minimal glucose medium. By using the same gel-independent methods, Champion *et. al.* (2003), studied the proteomes of exponential and stationary phase *E. coli* grown in MOPS minimal glucose medium. In total for both growth phases, 520 different proteins were identified (Champion et al. in preparation). Using 2D LC-MS/MS, 404 high quality proteins, which have more than one peptide identified, were identified by Corbin *et. al.* (2003). Using 2D PAGE/MS, 570 exponential phase proteins were identified by Project CyberCell (Sundararaj *et al.*, 2004). Overall 1047 proteins have been identified of the more than 4000 proteins that are predicted from the *E. coli* genome.

About 75% predicted proteins haven't been identified by current proteomic studies. One reason is that not all gene products are expressed under the same conditions. Expression is regulated to adapt to different growth conditions, such as different growth media and different growth phases. Aanother reason is that global proteomics are biased towards high abundance proteins (Champion *et al.*, 2003; Corbin *et al.*, 2003; Tonella *et al.*, 2001). The codon adaptation index (CAI) is a parameter used to predict the expression level of genes (Eyre-Walker, 1996; Sharp and Li, 1987). Proteins with higher

CAI values (≥0.5) are predicted to be high abundance proteins, while proteins with low CAI values (<0.5) are predicted to be low abundance proteins. For proteins with a CAI greater than or equal to 0.5 of the whole proteome, 70% have been identified in the global proteomic studies just cited. For proteins with a CAI less than 0.5, only 20% have been identified. One way to solve this problem is to apply methods with higher sensitivity. Another way is to apply sub-proteomics, which can enrich and identify low abundance proteins by focusing on specific groups of proteins.

Different physical properties have been studied using proteomic approach. Using tandem affinity purification (TAP) (Butland *et al.*, 2005) or his-tag purification (Arifuzzaman *et al.*, 2006), thousands of protein pair candidates were identified for *E. coli*. Protein carbonylation has been studies for cells under different stresses (Ballesteros *et al.*, 2001; Dukan and Nystrom, 1998). During stationary phase, several TCA cycle enzymes, glutamine synthetase, glutamate synthase, pyruvate kinase, DnaK, and H-NS are carbonylated (Dukan and Nystrom, 1998).

Methods to manage and display proteomic data

Studies of proteomics, genomics, transcriptomics, metabolomics, and other 'OMICS' create large amounts of data, which are difficult to organize either manually or using a computer spreadsheet such as Excel. Relational Database Management Systems (RDBMS), which can be supported by MySQL® or Oracle®, are easier and more efficient tools to store, organize and analyze data (Baclawski *et al.*, 1993). Using a database, not only the final results, but also details of every step of experiments, such as materials and methods, can be stored. To analyze and mine the proteomic data, scripts can be written to query such databases. Moreover, these databases can be connected to website interfaces using Perl-CGI, PHP and HTML and the data can be displayed to multiple users.

There are many web-based databases for *E. coli* proteomic and genomic studies, some of them are listed in Table 1.1. Among these databases, only a few of them are for *E. coli* proteomic studies, such as Project CyberCell (Sundararaj *et al.*, 2004). Right now, no *E. coli* proteomic database contains proteomics data from other groups. Since more and more proteomic data obtained from different studies, it is important and useful to have a database containing *E. coli* proteomic data from different group. This summary of data in database will facilitate the comparison of different studies.

As the amount of proteomic data increases, database and computational techniques become more important for analyzing data of proteomics. Proteomic studies in this dissertation will use these tools to analyze data. Moreover, web-based databases will be constructed to facilitate the management and displaying the data.

Studies of this dissertation and content of each chapter

This dissertation studied differences between exponential growth and stationary phase of *E. coli* cells at the protein level, and applied bioinformatics tools to manage, display the data generated by proteomics, and to display the genome annotation of different *E. coli* strains.

Chapter II describes the study of *E. coli* heparin-binding proteins from both exponential growth and stationary phase cells. Chapter III describes the study of the whole proteome for *E. coli* cells from both exponential growth and stationary phase using anion-exchange columns (AIX), 2D PAGE and MS/MS. Chapter IV describes a web-based database, Experiments in *E. coli* Proteomics (EEP), which manages data generated by proteomic studies. Chapter V describes studies of *E. coli* bioinformatics, which applies different bioinformatic tools on studies of genomics and proteomics. Chapter VI describes conclusions of this dissertation.

Table 1.1. *E. coli* web-based databases.

Name	URL	References
ASAP	https://asap.ahabs.wisc.edu/asap/home.php	(Glasner *et al.*, 2003)
AlignACE	http://arep.med.harvard.edu/ecoli_matrices/	(Robison *et al.*, 1998)
ColiBASE	http://colibase.bham.ac.uk/	(Chaudhuri *et al.*, 2004)
Colibri	http://genolist.pasteur.fr/Colibri/	(Medigue *et al.*, 1993)
E. coli genome project	http://www.genome.wisc.edu/	
EcoCyc	http://ecocyc.org/	(Karp *et al.*, 1996; Karp *et al.*, 1997, 1998; Karp *et al.*, 2000; Karp *et al.*, 2002; Keseler *et al.*, 2005; Paley and Karp, 1996)
EcoGene	http://bmb.med.miami.edu/EcoGene/EcoWeb/	(Rudd, 2000)
EchoBase	http://ecoli-york.org/	(Misra *et al.*, 2005)
GenProtEC	http://genprotec.mbl.edu/	(Riley and Space, 1996; Riley, 1997; Serres *et al.*, 2004)
GenoBase	http://ecoli.naist.jp/	
PEC	http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp	
Project CyberCell	http://www.projectcybercell.ca/	(Sundararaj *et al.*, 2004)
RegulonDB	http://www.cifn.unam.mx/Computational_Genomics/regulondb/	(Huerta *et al.*, 1998; Salgado *et al.*, 1999; Salgado *et al.*, 2000; Salgado *et al.*, 2001; Salgado *et al.*, 2004; Salgado *et al.*, 2006)

CHAPTER II

HEPAROME

Background

There are more than 4300 predicted protein-coding genes for *E. coli* (Riley *et al.*, 2006). No method can identify all the proteins at one time. One reason is that not all of the proteins are expressed in all growth conditions. Expression is regulated to adapt to different growth conditions, such as different growth media or different growth phases. Another reason is that studies of the whole proteome by current methods are not sensitive enough to detect low abundance proteins. However multiple methods can be applied to study different sub-proteomes, which are specific groups of proteins that share some physical property, such as phosphorylation, glycosylation or ubiquitination (Ferguson and Smith, 2003; Larsen and Roepstorff, 2000; Zhang *et al.*, 2004).

In this chapter, we describe the use of heparin chromatography to enrich for heparin-binding proteins of *E. coli* cells from both exponential growth and stationary phase cells[1]. Heparin is a negatively charged sulfated glycosaminoglycan (Fig. 2.1), which can bind to proteins such as certain receptors, growth hormones and enzymes involved in nucleic acid metabolism (Imai *et al.*, 2002). Heparin columns have been applied to proteomics of *Haemophilus influenzae* (Fountoulakis and Takacs, 1998; Langen *et al.*, 2000) and

---

[1]In this chapter, protein samples and SDS-PAGE were prepared by me. Mass spectrometry was run by Dr. Sam Perkins from Dr. David Russell's laboratory.

Mycoplasma pneumoniae (Ueberle *et al.*, 2002).

Because heparin is negatively charged, polyanion-binding proteins, such as nucleic acid-binding proteins are expected to be identified in the heparin binding sub-proteome, also called the "Heparome". By comparing the differences of the *E. coli* Heparome between exponential growth and stationary phase, proteins that are important during stationary phase for regulation, information transfer, transportation, and metabolism were expected to be identified.

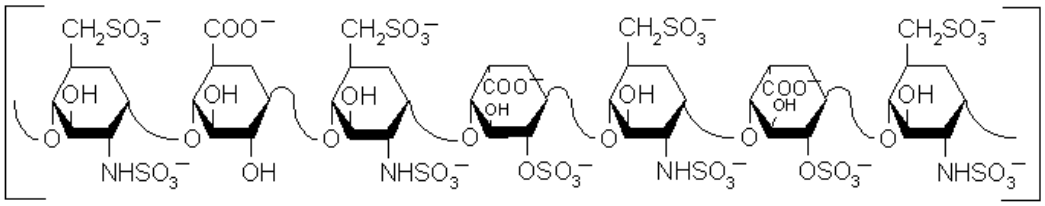


Figure 2.1. The structure of heparin, a sulfated poly-glycosaminoglycan.

Results

Identification of E. coli *heparin-binding proteins*

Exponential growth and stationary phase cells of *E. coli* K-12 strain MG1655 were grown and harvested as described in Materials and Methods. The strategy to identify heparin-binding proteins is shown in Figure 2.2. Clarified lysates were loaded onto heparin affinity columns as described in Materials and Methods. The majority of proteins were in the flow-through and about 10% of the total proteins (by Bradford assay) bound to the heparin column. Bound proteins were eluted with a salt gradient and collected in 16 fractions (Fig. 2.3). Proteins in each fraction were identified by MALDI-MS analysis of tryptic peptides combined with Applied Biosystems GPS Explorer Software. Two methods were used to generate the tryptic peptides. An aliquot of each fraction was digested with trypsin followed by separation of peptides by liquid chromatography (LC). In addition, each fraction was separated by SDS-PAGE (Fig. 2.4) followed by in-gel digestion of protein bands excised from the gel.

Many protein bands existed in adjacent fractions, and the patterns of protein bands in adjacent fractions were changing continuously. Observed molecular weights of protein bands ranged from 15 kD to 230 kD. Later fractions contained more protein bands with lower observed molecular weights.

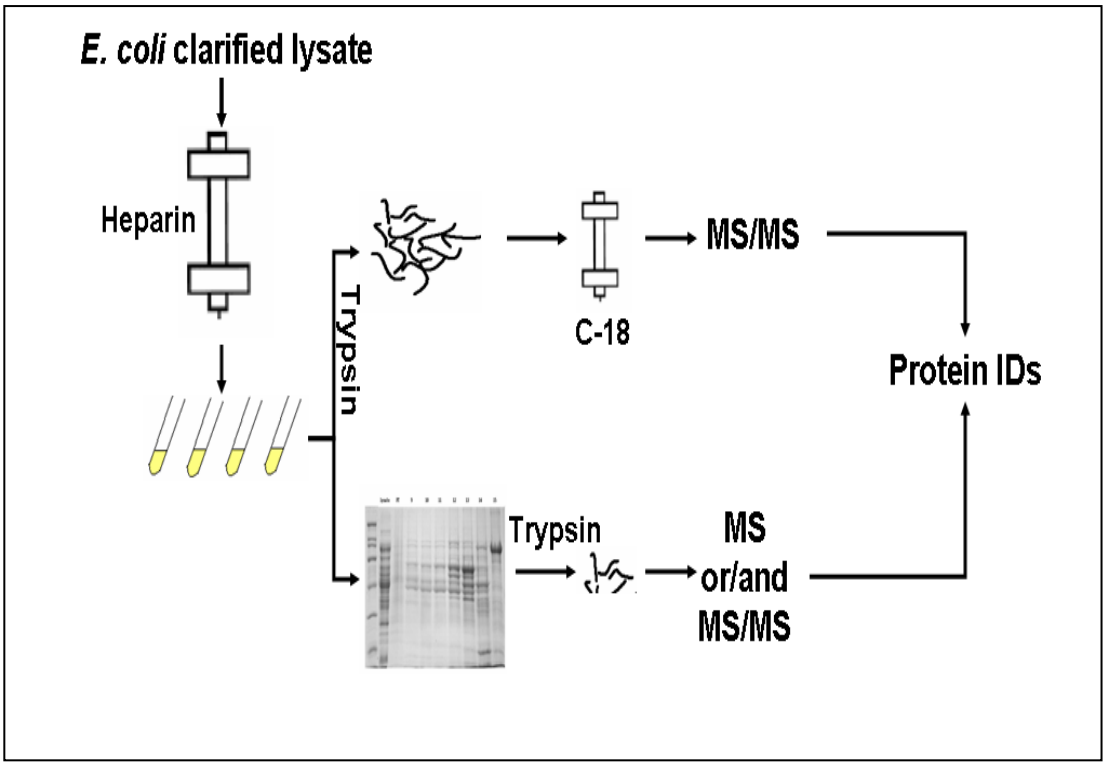
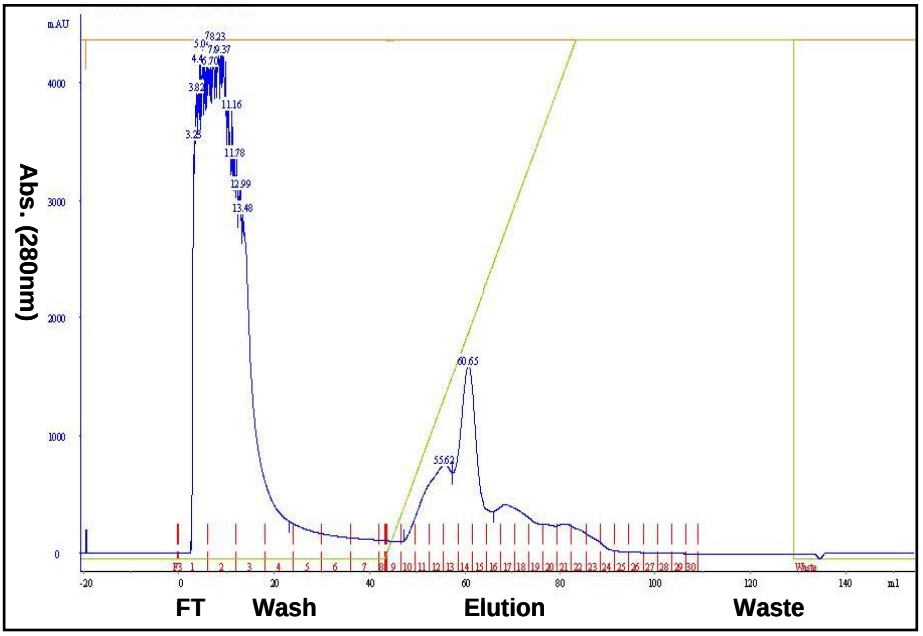


Figure 2.2. Flowchart of the strategy to identify heparin-binding proteins of *E. coli* cells. Clarified cell lysates were loaded on heparin-HPLC columns. Heparin-binding proteins were collected in 16 fractions. An aliquot of each fraction was separated on SDS-PAGE. Proteins on gels were excised, in-gel digested with trypsin, and identified by MS or/and MS/MS. Another aliquot of each fraction was dialyzed, digested by trypsin, separated by LC and identified by MS/MS. Proteins identified multiple times were counted only once.

A)



B)

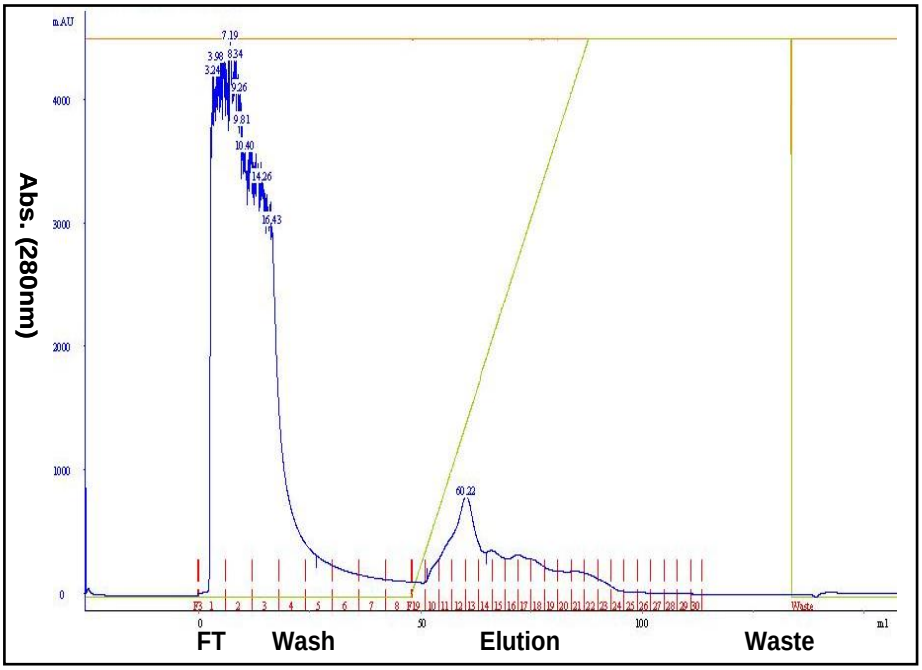
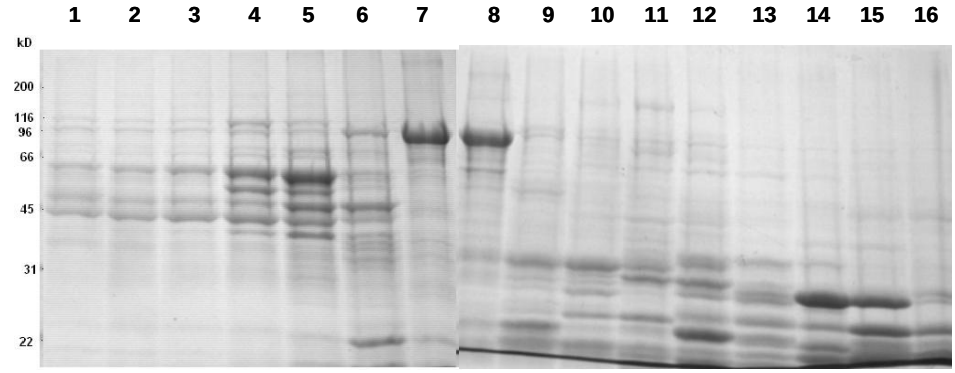


Figure 2.3. UV_{280} traces from the AKTA HPLC heparin column for samples from exponential growth (A) and stationary phase (B) of *E. coli* cells. The dark line is the UV_{280} plot in absorbance units and the grey line is the salt gradient as a percent.

A)



B)

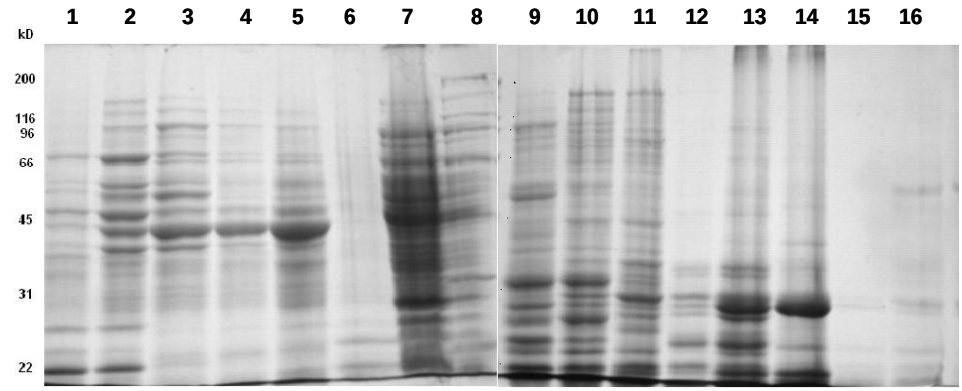


Figure 2.4. SDS-PAGE separation for heparin columns fractions from exponential growth (A) and stationary phase (B) lysate of *E. coli* cells.

The number of non-redundant proteins identified in each fraction ranged from 21 to 124 (Fig. 2.5). In total 1,358 redundant proteins were identified by either in-gel/MS/MS or LC/MS/MS from these 32 elution fractions (see table in APPENDEX A). Most of the proteins were identified in multiple fractions and/or in both exponential growth and stationary phase samples. For example, the product of *rpoA* was found in fractions 3-14 and 2-11 from the exponential growth and stationary phase samples, respectively.

318 non-redundant proteins were identified as heparin-binding proteins (Table 2.1A). 240 non-redundant proteins were identified from the exponential growth sample (Table 2.1B), 234 non-redundant proteins were identified from the stationary phase sample (Table 2.1C), and 156 non-redundant proteins were found in both growth phases (Table 2.1A and Table 2.1B) (Fig. 2.6).

The theoretical molecular weights of the 318 heparin-binding proteins ranged from 6 kD to 170 kD, and the molecular weight distribution of heparin-binding proteins was similar to that of the whole proteome (Fig. 2.7A). Since heparin is a highly sulfated glycosaminoglycan, we expected mostly basic proteins to bind to the heparin column. However theoretical pIs of the heparin binding proteins ranged from 4.5 to 11.5, and the pI distribution of heparin-binding proteins was similar to that of predicted proteins from the whole genome of *E. coli* (Fig. 2.7B). Based on this fact, we concluded that protein binding to heparin might be due to the charges on patches of proteins (Sendak and Bensadoun, 1998) instead of the overall pI.

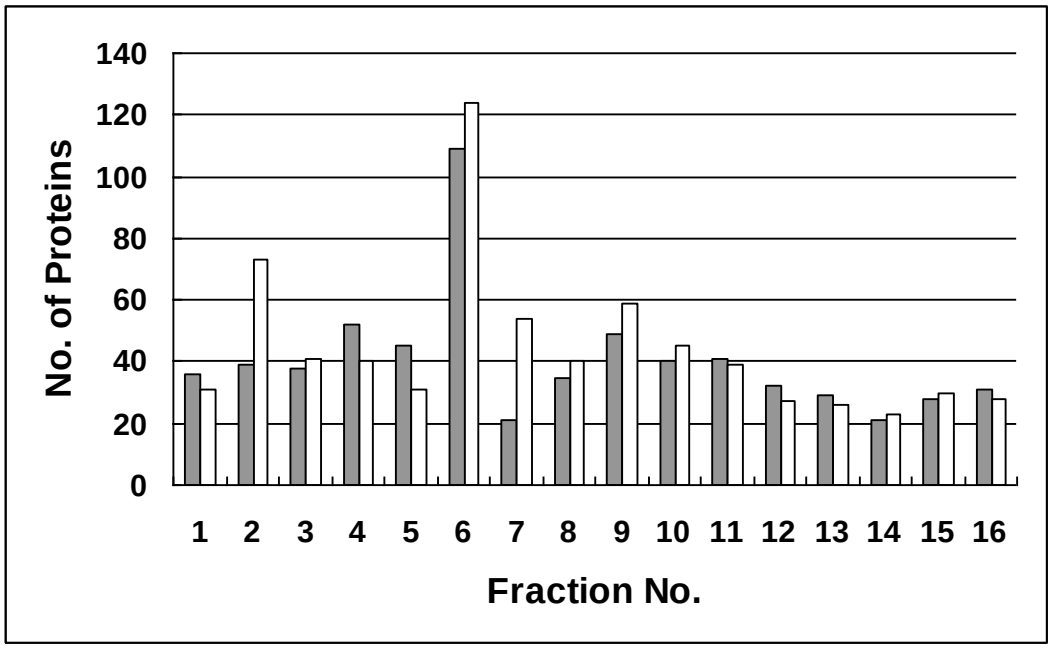


Figure 2.5. Distribution of protein identification on heparin columns. The number of non-redundant proteins identified in each fraction was plotted for the exponential growth sample (gray) and the stationary phase sample (white).

Table 2.1A. Heparin-binding proteins identified from exponential growth and stationary phase *E. coli* cells.

	In-gel/MS	**LC/MS**	**Combined**
Exponential phase			
Redundant proteins	471	489	960
Non-redundant proteins	151	178	240
Stationary phase			
Redundant proteins	358	604	962
Non-redundant proteins	117	196	234
Totals			
Redundant proteins	829	1093	1922
Non-redundant proteins	195	241	318

Table 2.1B. Heparin-binding proteins found in exponential growth of *E. coli* cells. Gene name and SwissProt ID are displayed for every protein. Entries bolded are proteins identified only in the exponential growth phase.

accA/P30867, accC/P24182, accD/P08193, aceE/P06958, adhE/P17547, **alaS/P00957**, arcA/P03026, **argA/P08205**, aroC/P12008, **aroH/P00887**, atpA/P00822, atpD/P00824, **bglX/P33363**, **ccmH/P33925**, **creA/P08367**, **crl/P24251**, crp/P03020, **cysH/P17854**, **dapE/P24176**, **deaD/P23304**, def/P27251, **dinG/P27296**, dksA/P18274, dnaK/P04475, **era/P06616**, erfK/P39176, fabA/P18391, fabB/P14926, **fabF/P39435**, fabG/P25716, fabH/P24249, fabI/P29132, **fabR/P27307**, **fadR/P09371**, ftnA/P23887, **fur/P06975**, **gapA/P06977**, ghrA/P75913, **gidB/P17113**, **glf/P37747**, glgB/P07762, **glgC/P00584**, **glmU/P17114**, **glnE/P30870**, glnS/P00962, gltB/P09831, gltX/P04805, **glyA/P00477**, gnd/P00350, guaB/P06981, gyrA/P09097, **gyrB/P06982**, hflX /P25519, **hisG/P10366**, hisS/P04804, hns/P08936, **hscC/P77319**, hupA/P02342, hupB/P02341, **hycE/P16431**, ihfA/P06984, ihfB/P08756, infA/P02998, infB/P02995, infC/P02999, iscS/P39171, kdgR/P76268, **lexA/P03033**, ligA/P15042, lon/P08177, lpd/P00391, **lpxB/P10441**, lrp/P19494, malP/P00490, map/P07906, mdoG/P33136, **metE/P25665**, metF/P00394, **metG/P00959**, metJ/P08338, mfd/P30958, **miaA/P16384**, **minE/P18198**, **mnmG/P17112**, moaB/P30746, **moaC/P30747**, mprA/P24201, mraW/P18595, **mug/P43342**, **nuoG/P33602**, **nusB/P04381**, **ompA/P02934**, ompC/P06996, **ompF/P02931**, ompR/P03025, **ompT/P09169**, oppA/P23843, oxyR/P11721, **panC/P31663**, parC/P20082, **parE/P20083**, **pbpC/P76577**, pcnB/P13685, pdxH/P28225, pepA/P11648, pgk/P11665, **pheS/P08312**, **pheT/P07395**, **pnp/P05055**, **pntA/P07001**, ppc/P00864, ppk/P28688, prc/P23865, **proC/P00373**, purU/P37051, rbfA/P09170, **rcsB/P14374**, relA/P11585, **rhlB/P24229**, rhlE/P25888, rho/P03002, **rlpB/P10101**, rluC/P23851, **rluE/P75966**, **rne/P21513**, rnr/P21499, rob/P27292, rplA/P02384, rplB/P02387, rplC/P02386, rplD/P02388, rplE/P02389, rplF/P02390, rplI/P02418, rplJ/P02408, rplK/P02409, rplL/P02392, rplM/P02410, rplN/P02411, rplO/P02413, rplP/P02414, rplQ/P02416, rplR/P02419, rplS/P02420, rplT/P02421, rplU/P02422, rplV/P02423, rplW/P02424, rplX/P02425, rplY/P02426, rpmA/P02427, rpmB/P02428, rpmC/P02429, rpoA/P00574, rpoB/P00575, rpoC/P00577, **rpoD/P00579**, rpsA/P02349, rpsB/P02351, rpsC/P02352, rpsD/P02354, rpsE/P02356, rpsF/P02358, rpsG/P02359, rpsH/P02361, rpsI/P02363, rpsJ/P02364, rpsK/P02366, rpsM/P02369, rpsN/P02370, rpsO/P02371, rpsP/P02372, rpsQ/P02373, rpsR/P02374, rpsS/P02375, rpsT/P02378, rpsU/P02379, **rsd/P31690**, rsmC/P39406, **rsuA/P33918,** secA/P10408, **selB/P14081**, skp/P11457, slyA/P55740, **spoT/P17580**, stpA/P30017, sucA/P07015, sucB/P07016, suhB/P22783, talB/P30148, **tgt/P19675**, thrS/P00955, **tktA/P27302**, **trpR/P03032**, trpS/P00954, **ttk/P06969**, tufA/tufB/P02990, **tyrR/P07604**, tyrS/P00951, **ugpQ/P10908**, **uidR/Q59431**, **ung/P12295**, uvrB/P07025, **uvrY/P07027**, xthA/P09030, yaaA/P11288, yaeH/P37048, **yafC/P30864**, **yafJ/Q47147**, **yafL/Q47151**, ybaK/P37175, ybeZ/P77349, **ybgK/P75745**, ybiB/P30177, **ycbB/P22525**, ycgK/P76002, **yciK/P31808**, ycjX/P76046, **ydcC/P28917**, **ydfH/P77577**, ydgH/P76177, ydjA/P24250, yegQ/P76403, **yejK/P33920**, **yfaA/P17994**, **yfhD/P30135**, **yfiF/P33635**, **yfjH/P52123**, yggH/P32049, yggS/P52054, yhaJ/P42623, yhhX/P46853, **yibK/P33899**, **yibQ/P37691**, yihA/P24253, **yjeQ/P39286**, yjgA/P26650, ykgM/P71302, ynhG/P76193, zwf/P22992

Table 2.1C. Heparin-binding proteins found in stationary phase of *E. coli* cells. Gene name and SwissProt ID are displayed for every protein. Entries bolded are proteins identified only in the stationary phase.

accA/P30867, accC/P24182, accD/P08193, **aceA/P05313**, aceE/P06958, **aceF/P06959**, adhE/P17547, **ahpC/P26427**, **aidB/P33224**, **allR/P77734**, arcA/P03026, aroC/P12008, atpA/P00822, atpD/P00824, **atpF/P00859**, **bfr/P11056**, **carB/P00968**, **cbpA/P36659**, **clpA/P15716**, **cpsB/P24174**, crp/P03020, **dacB/P24228**, def/P27251, **dinB/Q47155**, dksA/P18274, dnaK/P04475, **dnaX/P06710**, **dps/P27430**, **eda/P10177**, **envC/P37690**, erfK/P39176, fabA/P18391, fabB/P14926, fabG/P25716, fabH/P24249, fabI/P29132, **fmt/P23882**, ftnA/P23887, **gadB/P28302**, ghrA/P75913, glgB/P07762, **glnK/P38504**, glnS/P00962, gltB/P09831, gltX/P04805, gnd/P00350, **gpmA/P31217**, **groL/P06139**, **grxB/P39811**, guaB/P06981, **gutQ/P17115**, gyrA/P09097, hflX/P25519, **hisB/P06987**, hisS/P04804, hns/P08936, **htpG/P10413**, hupA/P02342, hupB/P02341, **icd/P08200**, ihfA/P06984, ihfB/P08756, **ilvE/P00510**, infA/P02998, infB/P02995, infC/P02999, iscS/P39171, **ivy/P45502**, kdgR/P76268, **lacZ/P00722**, ligA/P15042, lon/P08177, lpd/P00391, **lpxD/P21645**, lrp/P19494, malP/P00490, **malQ/P15977**, **manX/P08186**, map/P07906, **mdh/P06994**, mdoG/P33136, metE/P25665, metJ/P08338, **metQ/P28635**, mfd/P30958, **mglA/P23199**, moaB/P30746, mprA/P24201, mraW/P18595, **mukB/P22523**, **nadR/P27278**, **nagA/P15300**, **nikE/P33594**, **nusG/P16921**, ompC/P06996, ompR/P03025, oppA/P23843, oxyR/P11721, parC/P20082, pcnB/P13685, pdxH/P28225, pepA/P11648, pgk/P11665, **phoB/P08402**, **polA/P00582**, ppc/P00864, ppk/P28688, **ppx/P29014**, prc/P23865, **purC/P21155**, **purR/P15039**, purU/P37051, rbfA/P09170, **rbsA/P04983**, relA/P11585, **relE/P07008**, rhlE/P25888, rho/P03002, **rlmB/P39290,** rluC/P23851, **rmf/P22986**, **rnc/P05797**, rnr/P21499, rob/P27292, rplA/P02384, rplB/P02387, rplC/P02386, rplD/P02388, rplE/P02389, rplF/P02390, rplI/P02418, rplJ/P02408, rplK/P02409, rplL/P02392, rplM/P02410, rplN/P02411, rplO/P02413, rplP/P02414, rplQ/P02416, rplR/P02419, rplS/P02420, rplT/P02421, rplU/P02422, rplV/P02423, rplW/P02424, rplX/P02425, rplY/P02426, rpmA/P02427, rpmB/P02428, rpmC/P02429, **rpmD/P02430**, rpoA/P00574, rpoB/P00575, rpoC/P00577, rpsA/P02349, rpsB/P02351, rpsC/P02352, rpsD/P02354, rpsE/P02356, rpsF/P02358, rpsG/P02359, rpsH/P02361, rpsI/P02363, rpsJ/P02364, rpsK/P02366, rpsM/P02369, rpsN/P02370, rpsO/P02371, rpsP/P02372, rpsQ/P02373, rpsR/P02374, rpsS/P02375, rpsT/P02378, rpsU/P02379, rsmC/P39406, secA/P10408, **selA/P23328**, **serA/P08328**, skp/P11457, slyA/P55740, **sodB/P09157**, **speG/P37354**, **spy/P77754**, **sthA/P27306**, stpA/P30017, sucA/P07015, sucB/P07016, suhB/P22783, talB/P30148, thrS/P00955, **tig/P22257**, **tpx/P37901**, trpS/P00954, **truA/P07649**, **truD/Q57261**, **tsf/P02997**, tufA/tufB/P02990, tyrS/P00951, uvrB/P07025, **wcaI/P32057**, **wrbA/P30849**, xthA/P09030, yaaA/P11288, **yaeB/P28634**, yaeH/P37048, **yaiN/P55756**, ybaK/P37175, **ybeX/P77392**, ybeZ/P77349, ybiB/P30177, ycgK/P76002, **yciH/P08245**, yciK/P31808, **ydcP/P76104**, ydgH/P76177, ydjA/P24250, **yeaO/P76243**, yegQ/P76403, **yffB/P24178**, **ygeA/P03813**, yggH/P32049, yggS/P52054, yhaJ/P42623, **yhbJ/P33995**, **yhhK/P37613**, yhhX/P46853, **yicC/P23839**, yihA/P24253, **yihI/P32130**, yjgA/P26650, ykgM/P71302, ynhG/P76193, **yraL/P45528**, zwf/P22992

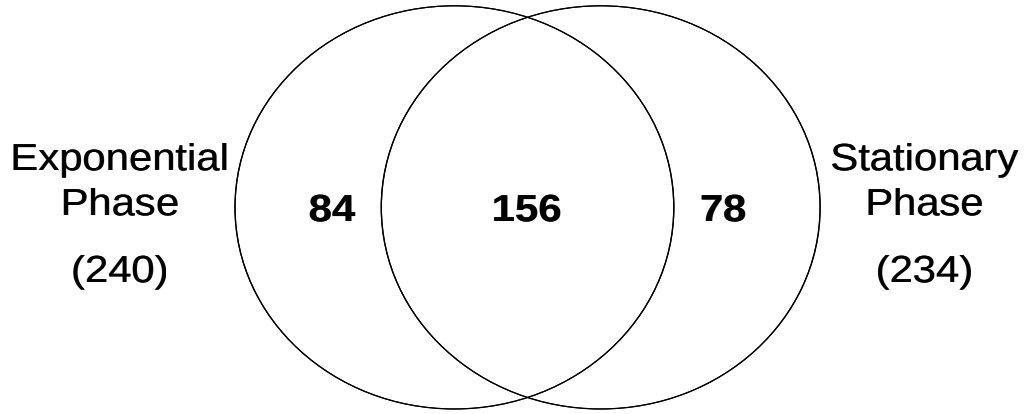
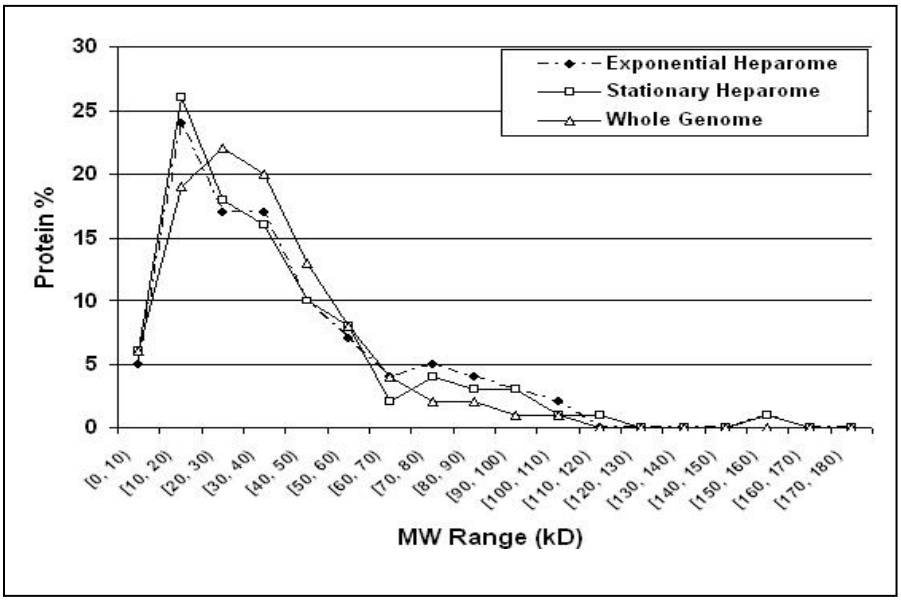


Figure 2.6. Comparison of heparin-binding proteins between exponential growth and stationary phase. 240 proteins were found in exponential growth, 234 were found for stationary phase, and 156 were found in both. In total, 318 proteins were identified as heparin-binding proteins.

A)



B)

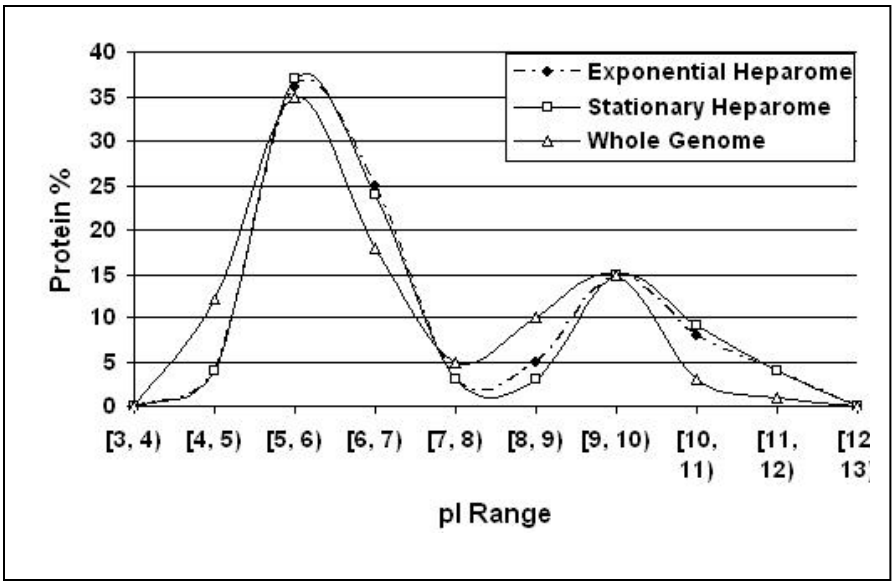
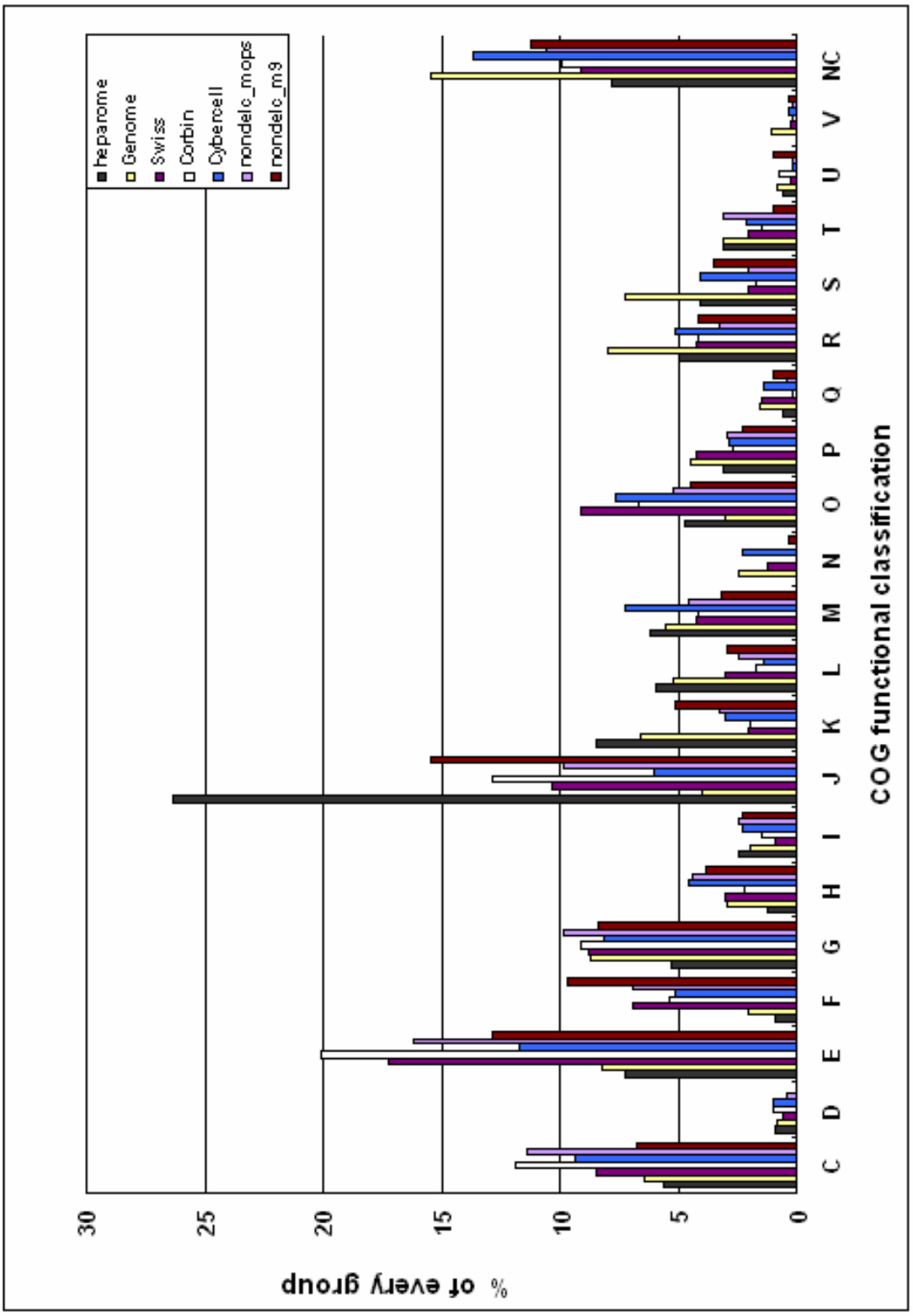


Figure 2.7. Comparisons of the distributions of molecular weight (A) and pI (B) between heparin-binding proteins and predicted proteins of the whole genome.

Heparin-binding proteins were found in all functional classifications (Fig. 2.8), as defined by the cluster of orthologus groups (COG) (Tatusov et al., 2003). 19 heparin-binding proteins were in the category of 'replication, recombination and repair'. 84 heparin-binding proteins were in the 'translation' category, including 47 ribosomal proteins, 9 tRNA synthetases, 3 protein chain elongation factors, and 3 protein chain initiation factors. 27 heparin-binding proteins were in the 'transcription' category, including 14 transcription regulators, 4 subunits of RNA polymerases, and RNase III. Overall 40.8% (130) of the heparin-binding proteins were in categories related to DNA- or RNA-binding proteins. The rest of them were distributed among other categories.

Comparing with other *E. coli* proteomic studies and proteins predicted from the whole *E. coli* genome, Heparome has high percentage in 'translation', 'transcription', 'replication, recombination and repair' categories (Fig. 2.8).

Figure 2.8. Functional classification by COG for heparin-binding proteins, Genome, Swiss 2D-PAGE, Corbin's, Project CyberCell, NonDeLC (M9) and NonDeLC (MOPS). C: Energy production and conversion; D: Cell cycle control, mitosis and meiosis; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking and secretion; V: Defense mechanisms; NC: Not in COGs.


heparome
Genome
Swiss
Corbin
Cybercell
nondelc_mops
nondelc_m3
30
25
20
15
10
5
0
% of every group
C D E F G H I J K L M N O P Q R S T U V NC
COG functional classification

Comparison between exponential growth and stationary phase

After the initial identification, comparisons were made of presence, predicted abundance, and chromatographic behavior between exponential growth and stationary phase.

Presence

For the 318 heparin-binding proteins, 84 of them were only found in exponential phase, and 78 of them were only found in stationary phase. By comparing proteins unique to exponential and stationary phase, two functional groups, transcription and cell wall/membrane biogenesis have more proteins for exponential phase (Fig. 2.9). For stationary phase unique Heparome, proteins were found more in three functional groups. They are carbohydrate transport and metabolism, post-translational modification, protein turnover, chaperones, and inorganic ion transport and metabolism (Fig. 2.9).

Abundance

To find heparin-binding proteins whose abundance changed between exponential growth and stationary phase, we used CyDye DIGE 2D PAGE and DeCyder™ v6.5 differential analysis software. Aliquots of every exponential or stationary phase elution fraction were pooled, labeled with fluorescent dyes, and separated by 2D-PAGE. Five heparin-binding were more abundant in stationary phase and 9 were more abundant in exponential growth phase (Table 2.2).The 5 more abundant heparin-binding proteins in stationary phase are the products of the *dps*, *elaB*, *ivy*, *manX* and *metJ* genes. The 9 more abundant heparin-binding proteins in exponential growth are the products of the *cysH*,

engB, *mdoG*, *metE*, *rplD*, *rplI*, *rpsB*, *rpsF*, and *rsuA* genes.

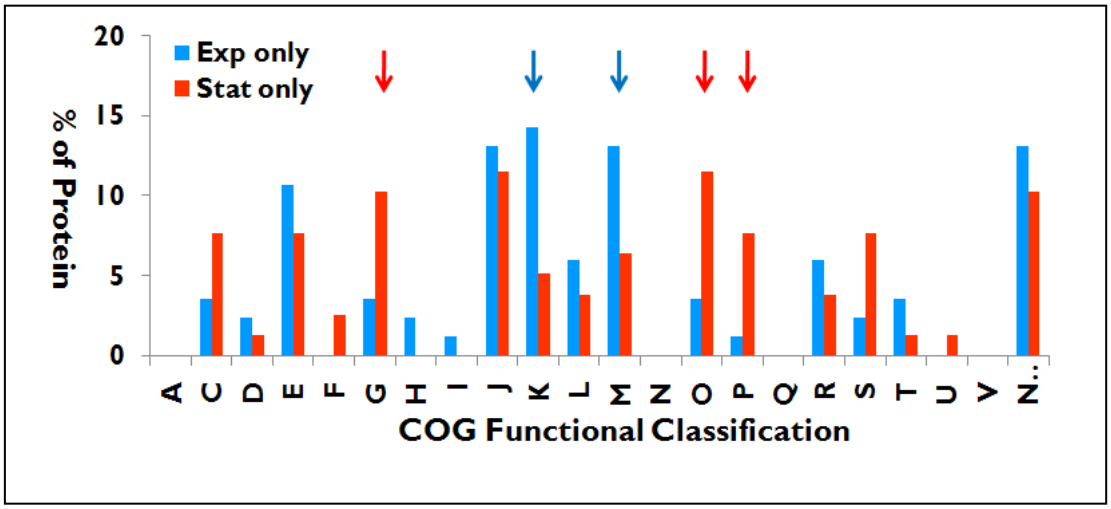


Figure 2.9. COG functional classification of heparin-binding proteins found only in exponential or stationary phase. G: Carbohydrate transport and metabolism. O: Post-translational modification, protein turnover, chaperones. P: Inorganic ion transport and metabolism. K: Transcription. M: Cell Wall/membrane biogenesis.

Table 2.2. Heparin-binding proteins with abundance changes between exponential growth and stationary phase

More abundant in stationary phase			
sp_id[a]	**gn**[b]	**b_number**[c]	**product**
P27430	*dps*	b0812	Stress response DNA-binding protein
P52084	*elaB*	b2266	Hypothetical protein
P45502	*ivy*	b0220	Inhibitor of vertebrate C-lysozyme
P08186	*manX*	b1817	Mannose phosphotransferase system, EIIAB component
P08338	*metJ*	b3938	Met regulon regulatory protein

More abundant in exponential phase			
sp_id	**gn**	**b_number**	**product**
P17854	*cysH*	b2762	Phosphoadenylyl sulfate (PAPS) reductase
P24253	*engB*	b3865	GTPase essential for cell cycle
P33136	*mdoG*	b1048	Periplasmic oligosaccharide synthesis
P25665	*metE*	b3829	Methionine synthase, cobalamin-independent; 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase; binds Zn(II)
P02388	*rplD*	b3319	50S ribosomal subunit protein L4
P02418	*rplI*	b4203	50S ribosomal subunit protein L9
P02351	*rpsB*	b0169	30S ribosomal subunit protein S2
P02358	*rpsF*	b4200	30S ribosomal subunit protein S6
P33918	*rsuA*	b2183	16S RNA pseudouridine 516 synthase

a: sp_id: SwissProt gene ID
b: gn: Gene Name
c: b_number: Blattner Number

Chromatography behavior---elution shifts

The elution profiles were compared for heparin-binding proteins present in both exponential growth and stationary phase. An elution shift was defined as a change in the peak elution position of ≥3 fractions. Among the 156 heparin-binding proteins found in both growth phases, 7 had an elution shift between these two growth phases (Table 2.3). These proteins are the products of the *def*, *dnaK*, *ftnA*, *ppc*, *ppk*, *relA* and *yaeH* genes. The elution positions of FtnA were verified by enhanced chemiluminescence (ECL) western (Fig. 2.10).

Validation of protein identification

Validation of proteins from in-gel/MS/MS--difference between Obs. MW and theoretical MW

To validate the protein identifications from in-gel/MS/MS, we compared the observed molecular weight (Obs. MW) on the gel with the theoretical molecular weight (MW). Among the 829 redundant proteins found by in-gel/MS/MS, 77% had observed molecular weights that were of the theoretical molecular weight (MW) (Fig. 2.11). The other proteins could be processed, degraded, or run aberrantly on SDS-PAGE.

Table 2.3. Proteins with elution shift on the heparin column between exponential growth and stationary phase

sp_id[a]	gn[b]	b_number[c]	Product	Fractions	
				Exp.[d]	Stat.[e]
P27251	*def*	b3287	peptide deformylase	5, 6	2
P04475	*dnaK*	b0014	chaperone Hsp70; DNA biosynthesis; autoregulated heat shock proteins	1	6, 7
P23887	*ftnA*	b1905	cytoplasmic ferritin (an iron storage protein)	1	6
P00864	*ppc*	b3956	phosphoenolpyruvate carboxylase	1	6
P28688	*ppk*	b2501	polyphosphate kinase	1	10
P11585	*relA*	b2784	(p)ppGpp synthetase I (GTP pyrophosphokinase); regulation of RNA synthesis; stringent factor	6	12, 14
P37048	*yaeH*	b0163	putative structural protein	10	6

a: sp_id: SwissProt gene ID
b: gn: Gene Name
c: b_number: Blattner Number
d: Exp.: Exponential phase
e: Stat.: Stationary phase

Figure 2.10. Identification of FtnA in heparin-column elution fractions by Western blotting with anti-FtnA antibody. Exp: exponential growth. Stat: stationary phase. CL: cell lysate. FT: flow through. 1 to 7 are elution fraction numbers.

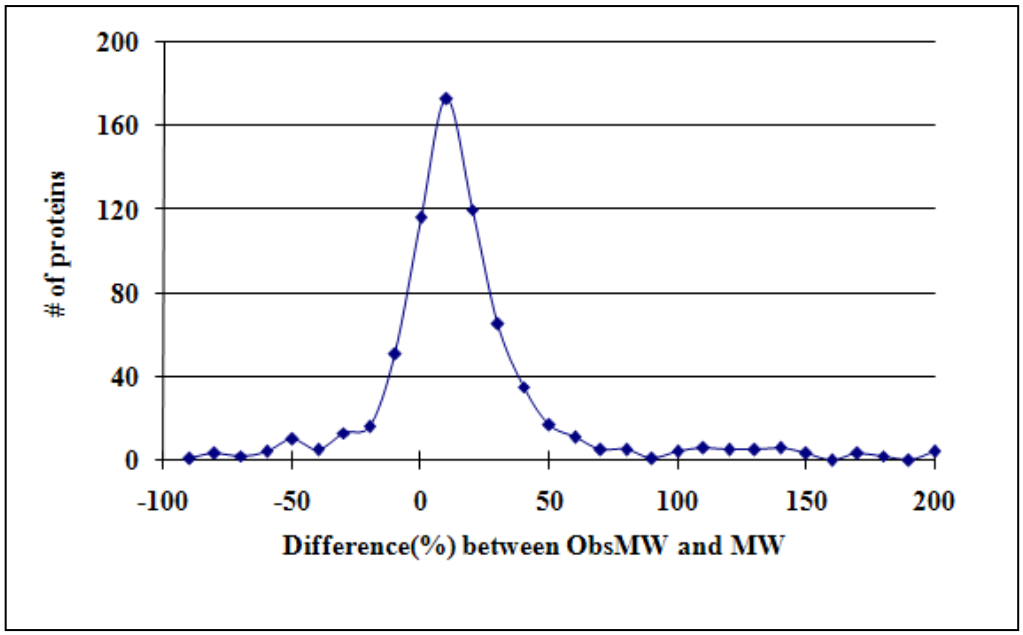


Figure 2.11. Distribution of difference between Obs. MW and theoretical MW for heparin-binding proteins identified by in-gel digestion combined with MS or/and MS/MS. Among 829 redundant proteins, 538 (77%) of them have the difference in ±30% between Obs. MW and theoretical MW.

Validation of protein identification from both in-gel/MS/MS and LC/MS/MS methods– pI distribution in every fraction

We predicted that proteins eluting in early fractions, which bind to heparin relatively loosely, would have a lower pI than proteins eluting in late fractions, which bind to the heparin tightly. To validate the identifications from both in-gel/MS/MS and LC/MS/MS, the pI distribution of proteins in every fraction was plotted (Fig. 2.12 and Fig. 2.13). The peak of the pI distributions shifted from 5 or 6 for proteins in early fractions to 10 or 11 for proteins eluting in later fractions. This result fit the property of heparin column, which bind basic protein more tightly.

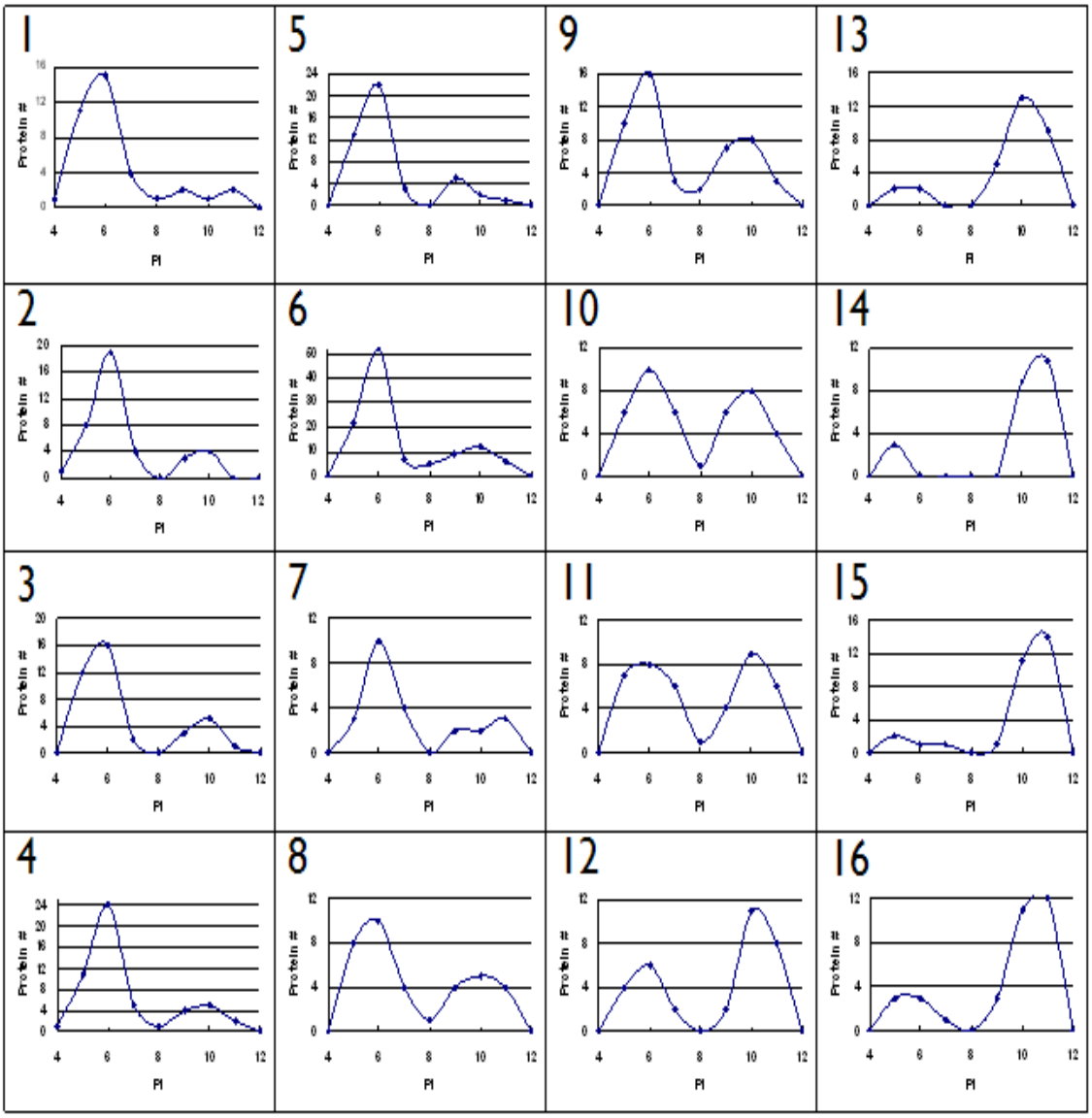


Figure 2.12. pI distribution of heparin-binding proteins in elution fractions of the exponential growth sample.

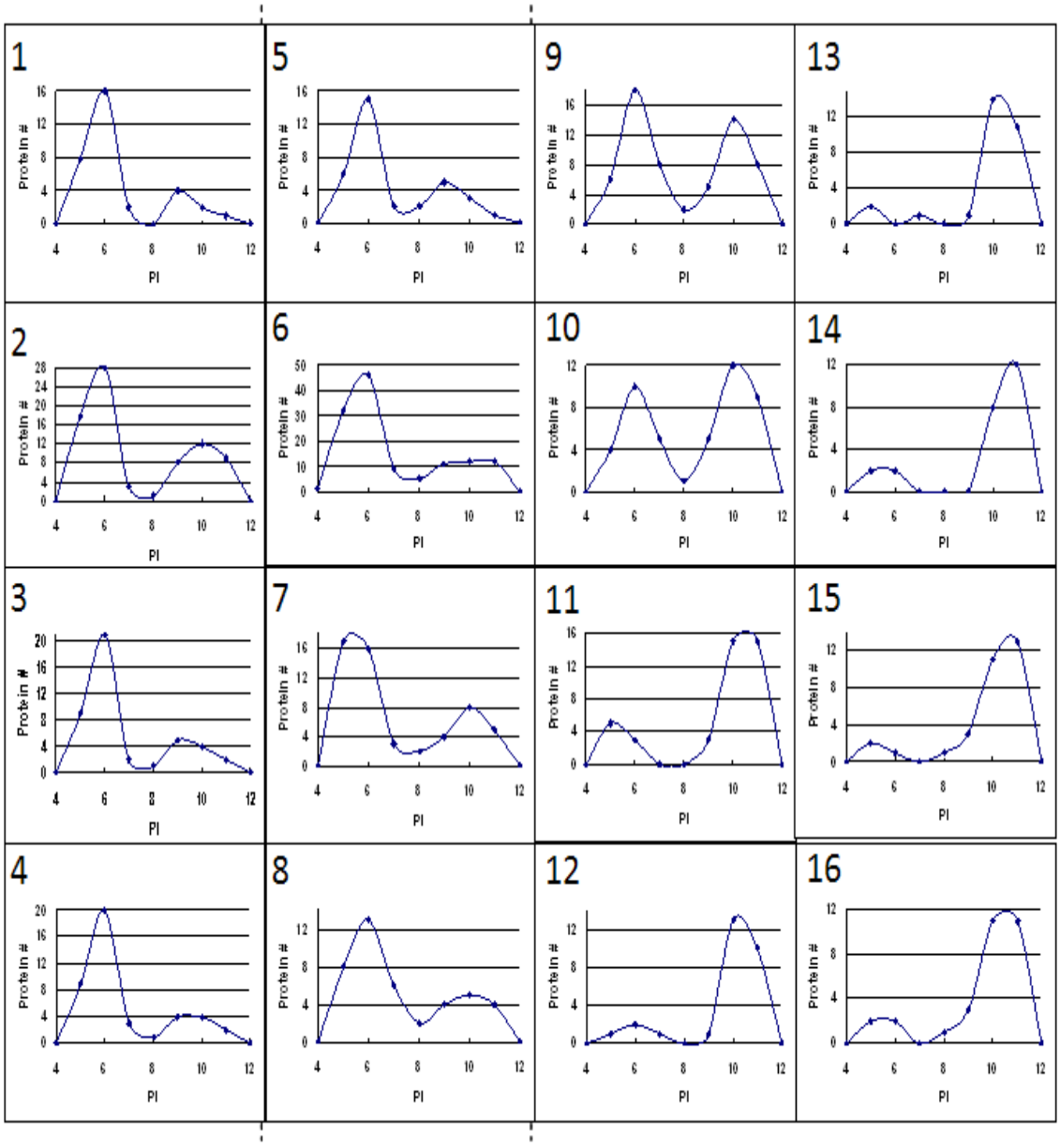


Figure 2.13. pI distribution of heparin-binding proteins in elution fractions of the stationary phase sample.

Discussion

In-gel/MS/MS and LC/MS/MS

In this study we used in-gel/MS/MS and LC/MS/MS to identify 195 and 241 heparin-binding proteins respectively, with 118 proteins identified by both methods.

Properties of heparin-binding proteins

The distribution of theoretical pI's of the heparin-binding proteins was similar to that of the predicted whole proteome of *E. coli*. We conclude that factor affecting proteins binding to the heparin column may include the charges on patches of proteins instead of the overall pI. Moreover, molecular weight distribution of heparin-binding proteins is also similar to that of whole genome.

Changes in stationary phase Heparome

Protein presence

For stationary phase Heparome, there are more proteins in carbohydrate transport and metabolism, post-translational modification, protein turnover, chaperones, and inorganic ion transport and metabolism. It is not surprised that *E. coli* cells need more chaperon proteins since more misfolding proteins are generated when cells enter stationary phase (Mujacic and Baneyx, 2007). Also due to depletion of nutrient, cells in stationary phase might also increase the expression of carbohydrate transport proteins to obtain different carbon sources.

Protein abundance

Among the 5 more abundant heparin-binding proteins in stationary phase, the

product of *dps* is a stress response DNA-binding protein that is known being up-regulated in stationary phase. The product of *ivy/ykfE* is the inhibitor of C-lysozyme (Deckers *et al.*, 2004). With increasing amount of C-lysozme, *E. coli* cells might increase the resistance of antibacterial lysozyme in stationary phase. The product of *manX* is one of the components of the major transporter for the mannose (Erni and Zanolari, 1985; Erni *et al.*, 1987) and the product of *metJ* is repressor of the methionine biosynthetic genes (Usuda and Kurahashi, 2005). These data suggest that in stationary phase of glucose starvation, *E. coli* cells increase the ability to transfer other sugar, such as mannose, and decrease the biosynthesis of methionine.

Chromatography behavior on heparin columns

Before behavior of proteins on heparin columns was examined, we asked if the amount will affect the behavior of proteins. By loading different amount of proteins onto same heparin column separately, elution profiles of heparin-binding proteins do not change (Fig. 2.14). Based on the properties of heparin column, we also predicted that most of the proteins that eluted early would be more acidic, while proteins eluted late would be more basic. The peaks of the protein's pI in early fractions are 5-6, while those of late fractions are 10-11, which agrees with our prediction. Most of the heparin-binding proteins with pIs lower than 5 were found in early fractions (1 to 8), such as the products of the *ahpC*, *bfr*, *dksA*, *dnaK*, *ftnA*, *gnd*, *groL*, *hscC*, *metQ*, *purC*, *rpoD*, *tpx* and *tig* genes. Most of the heparin-binding proteins with pIs higher than 10 were found in late fractions (9 to16th), such as products of *rhlE*, *rluE*, *rmf* and ribosomal proteins. The presence of proteins with high pIs in early fractions or proteins with low pIs in late

fractions might be caused by protein-protein interactions, protein and small molecule binding, changes of protein conformations, or post-translational modifications.

Proteins found with elution shift on heparin column belong to different biological processes, which indicate specific changes when cells enter stationary phase. The changes of behavior on the heparin column might be due to protein stability, changes in protein conformations, post-translational modifications (Dukan and Nystrom, 1998; Fredriksson *et al.*, 2005; Tamarit *et al.*, 1998), or protein-protein interactions.

For example, the product of *ftnA* is a cytoplasmic ferritin, which stores free iron during stationary phase (Andrews, 1998). In stationary phase, more ferritin is expressed to form a spherical protein which is composed of 24 identical subunits. This spherical protein contains a large hollow centre that acts as an iron-storage cavity to accommodate at least 2000 iron. Since FtnA is negatively charged, most of them flew through the heparin column. FtnA from exponential phase elutes out early, but FtnA from stationary phase binds to heparin column more tightly. Since heparin column is negatively charged, surface of ferritin from stationary phase could have more positively charges caused by post-translational modification or conformation changes. To test if the conformation changes, size exclusion column combined with ECL western could be applied to analyze the size of the ferritin in these two growth phases. Ferritin in monomer will be 19 kD and 24mer will be around 456 kD. Also oligomers between monomer and 24mer could

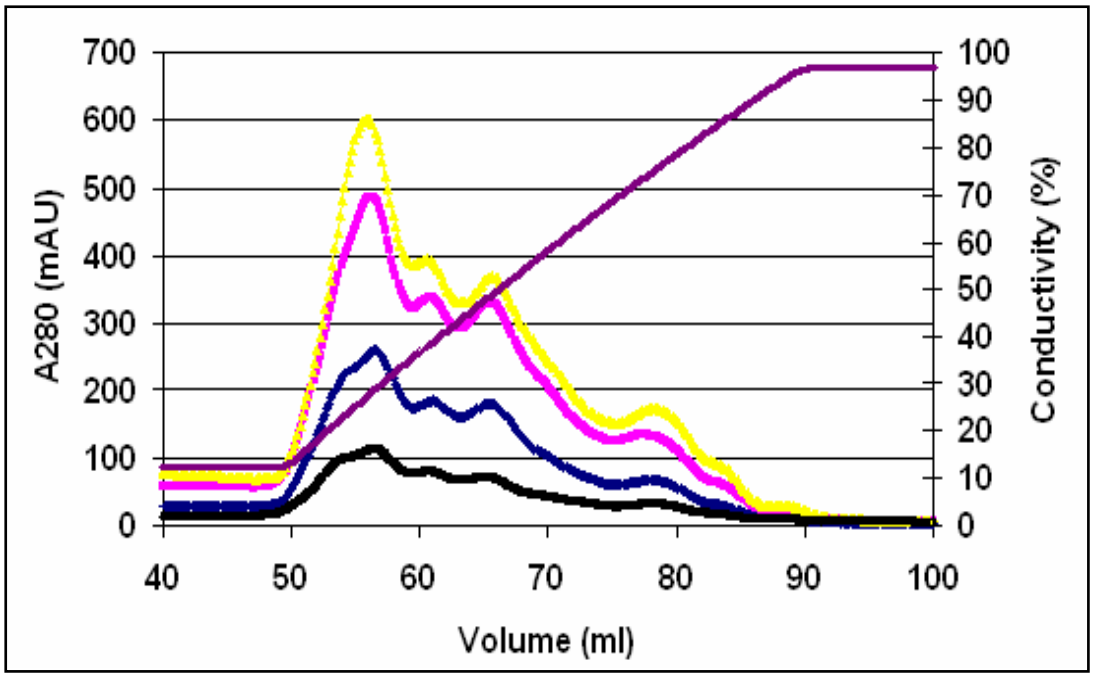
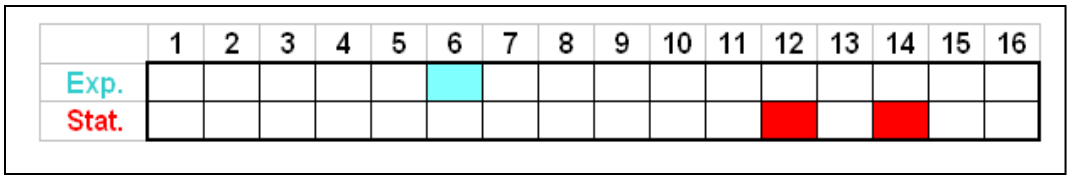


Figure 2.14. Elution profiles of different amount protein loaded on heparin columns. 10 mg (black), 20 mg (blue), 40 mg (pink), and 60 mg (yellow) proteins from stationary phase *E. coli* cells were loaded onto 5ml heparin HiTrap columns. Heparin-binding proteins were eluted with increasing NaCl concentration. Ratio for heights of four elution curves is about 1:2:4:4.8. Capacity of this column is about 50 mg proteins in cell lysate.

be detected. To test if it is due to the post-translational modification changes, 2D PAGE combined with ECL western or followed with MS and tandem MS could be applied. If the post-translational modification changes, different spots of FtnA protein could show up on 2D PAGE which could be detected by ECL Western. Also MS and tandem MS could differentiate same peptides with different modifications.

Another example is RelA, (p)ppGpp synthetase I and stringent factor. RelA from exponential phase eluted out in fraction 6^{th} from heparin column, while RelA from stationary phase eluted out later in fraction 12^{th} to 14^{th} (Fig. 2.15A). RelA binds to blocked ribosomes in stationary phase (Ramagopal and Davis, 1974; Wendrich *et al.*, 2002) and its activity is regulated by N-terminal of L11 (Yang and Ishiguro, 2001). Elution patterns of ribosomal proteins were different between exponential and stationary phases (Fig. 2.15B). We proposed that property changes of ribosomal protein affect the chromatography behavior of RelA from these two growth phase samples.

A)



B)

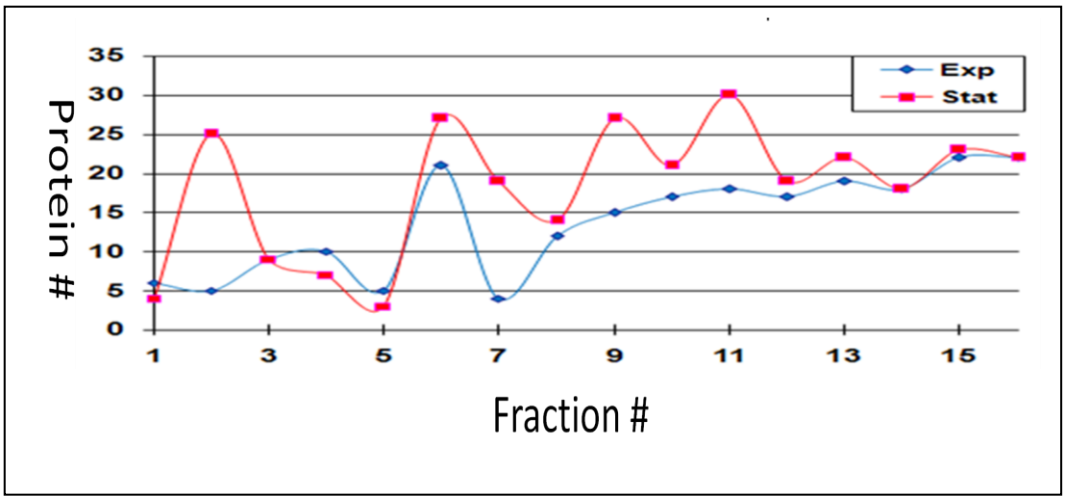


Figure 2.15. A) Elution positions of RelA on heparin columns. B) Distribution of ribosomal proteins on heparin columns. Exp. exponential phase. Stat. stationary phase.

Heparin-binding proteins and other proteomic studies

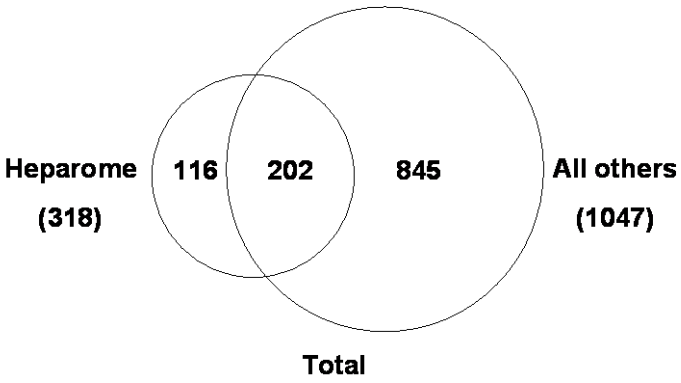
The Heparome includes proteins in all functional classifications. Compared with other proteomic studies (Champion *et al.*, 2003; Corbin *et al.*, 2003; Tonella *et al.*, 2001) (Project CyberCell; Champion, et al., 2005 submitted) and the whole genome, the Heparome enriched proteins in 3 categories: 'translation', 'transcription' and 'replication, recombination and repair'. The codon adaptivity index (CAI) (Eyre-Walker, 1996; Sharp and Li, 1987) values were applied to predict protein abundance. Proteins with high CAI values are predicted high abundance proteins, while proteins with low CAI values are predicted low abundance proteins. Most of the heparin-binding proteins in the 'translation' category are predicted high abundance proteins, while most of heparin-binding proteins in 'transcription' category are predicted low abundance proteins. There are 19 heparin-binding proteins that are in category of 'replication, recombination and repair'.

Among the 84 heparin-binding proteins in the 'translation' category, there are 47 ribosomal proteins, 9 tRNA synthetase, 3 protein chain elongation factors, 3 protein chain initiation factors and other proteins related to metabolism of tRNA, rRNA and mRNA. CAI values of 70 heparin-binding proteins in this category are higher than 0.4. 47 out of these 70 proteins are high abundance ribosomal proteins. Among the 27 heparin-binding proteins in 'transcription' category, there are 14 transcription regulators, 4 subunits of RNA polymerases, and RNase III. CAI values of 16 heparin-binding proteins in this category are lower than 0.4. 10 of these 16 proteins are low abundance transcriptional regulator. In total 130 heparin-binding proteins (40.8%) are in these 3

categories that are DNA-binding or RNA-binding proteins. The rest of them distribute in other categories.

Comparing our data with other known proteomic studies (Champion *et al.*, 2003; Corbin *et al.*, 2003; Tonella *et al.*, 2001) (Project CyberCell; Champion et al., 2006 submitted) the Heparome found 116 novel proteins (Fig. 2.16A). In summary, current proteomic studies identified 1163 gene products of *E. coli* K-12 (Fig. 2.16A), which cover 27% of all coded proteins on genome. Most of the predicted high abundant proteins, which have CAI values higher than 0.6, were already identified (Fig. 2.16B), while most of the predicted low abundant proteins, which have CAI values lower than 0.4, have not been previously found. The Heparome makes a contribution to identify low abundance proteins that cover many categories of functional classifications. For example, *E. coli* K-12 has 177 transcription regulators (Table 2.4), 165 of which are predicted low abundant proteins (CAI value lower than 0.4), previous studies identified 24 transcription regulators in total. The Heparome identified 17 transcription regulators, 13 of which are unique to the Heparome (Table 2.4).

A)



B)

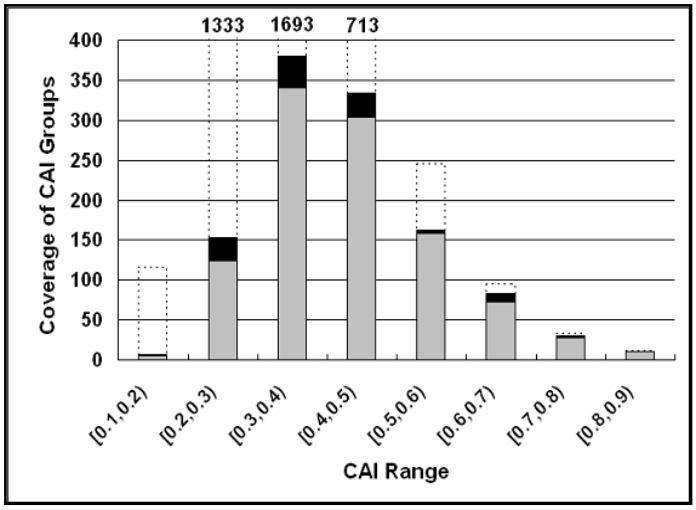


Figure 2.16. A) Venn diagram of summary for Heparome and other proteomic studies. (318 proteins were identified as heparin binding proteins and 1047 proteins were identified by 5 other proteomic studies (Tonella L, 2001; Corbin RW., 2003; Project CyberCell; Champion MM., 2003; and Champion MM., 2005 in press). 116 proteins were unique to Heparome. In total 1163 proteins have been identified for *E. coli* K12). B) Distribution of the CAI range of proteins identified by current proteomic studies. (116 newly identified heparin-binding proteins are presented in black, all other studies are presented in gray and dashed boxes are protein numbers of whole *E. coli* proteome in certain CAI range.)

Table 2.4. Genes of transcription regulators in *E. coli* K-12.

Transcriptional regulators							
abgR/ydaK	*fabR/yijC*	*lacI*	*nlp*	*yafC*	*ydcI*	*ygaV*	***yjhU***
acrR	***fadR***	*leuO*	*oxyR*	*yafY*	*ydcN*	*ygbI*	*yjiE*
agaR	*farR*	*lldR*	***pdhR***	*yagI*	*ydcR*	*ygeV*	*yjiR*
allR/ybbU	*fhlA*	*lrhA*	*perR*	*yagP*	*ydeW*	*ygfI*	*yjjM*
alpA	*fruR*	***lrp***	*phnF*	*yahB*	*ydfH*	*yggD*	*ynaK*
alsK	*fucR*	*lysR*	*prpR*	*yajF*	*ydhB*	*ygiP*	*yncC*
argR	***galR***	*malI*	*pspC*	***ybaD***	*ydhM*	*ygiT*	*yneJ*
arsR	***galS***	*malT*	*pspF*	*ybaO*	*ydjF*	*yhaJ*	*ynfL*
ascG	*gatR*	*marR*	***purR***	*ybbS*	***yeaT***	*yhcS*	*yphH*
asnC	*gcvA*	*metJ*	*rbsR*	*ybdM*	*yebK*	*yhfR*	*yqfE*
betI	*glcC*	***metR***	*rpiR*	*ybdO*	*yeeY*	*yhjC*	***yqgE***
cbl	***glpR***	*mhpR*	*sgcR*	***ybeF***	*yegW*	*yiaG*	*yqjI*
cueR	***gntR***	*mlc*	*slyA*	*ybhD*	*yehV*	*yiaJ*	*ytfA*
cynR	*hcaR*	*mntR/ybiQ*	*soxR*	*ybiH*	*yeiE*	*yiaU*	*ytfH*
cysB	*hdfR/yifD/yifA*	*mprA/emrR*	*srlR*	*ycaN*	*yfaH*	***yidP***	***ytfJ***
cytR	*hipB*	*mtlR*	*tdcA*	*ycdC*	*yfaX*	*yidZ*	*zntR*
deoR	*hyfR*	*nac*	*treR*	*ycfQ*	*yfeR*	*yieP*	
dgoR/yidW	*iciA*	*nadR*	*ttk*	*ycfX*	*yfeT*	*yihL*	
dicA	***iclR***	*nagC*	***tyrR***	*ycgE*	*yfhH*	*yihW*	
dsdC	***idnR***	*nanK*	*uidR*	***yciT***	*yfiE*	***yjdC***	
ebgR	*ilvY*	*nanR*	*uxuR*	*ycjC*	*yfjR*	*yjeB*	
envR	*iscR*	*nhaR*	*xapR*	*ycjW*	***ygaA***	*yjfQ*	
exuR	*kdgR*	*nikR*	*xylR_1*	*ycjZ*	*ygaE*	*yjhI*	

There are 177 transcription regulators. 24 of them (bold) were identified in previous proteomic studies (Tonella et al., 2001; Corbin et al., 2003; Project CyberCell; Champion et al., 2003; and Champion et al., 2006 submitted), 17 of them (underlined) were identified by Heparome.

Materials and methods

Sample preparation

E. coli K-12 strain MG1655 was inoculated from a single colony grown on minimal agar (0.2% glucose) into 5 ml of MOPS minimal glucose medium (0.4% glucose, 19 mM NH_4Cl, 1.32 mM K_2HPO_4, 2 μg/ml thiamine, 10 μg/ml uridine, 0.52 mM $MgCl_2$, 0.25 μM $CaCl_2$, 8.37 g/l MOPS, 0.72 g/l tricine, 48 mg/l K_2SO_4, 2.92 g/l NaCl, 3 mg/l $FeSO_4 \cdot 7H_2O$, and micronutrients) and grown overnight with aeration at 37°C for 16 to 18 hours. 5 ml of overnight culture was used to inoculate 2 liters of MOPS media, which was evenly divided into 2 separate 1-liter non-baffled flasks. Cultures were grown with aeration (250 RPM) in a 37°C water bath. 1 liter of the culture was harvested at mid-exponential growth ($OD_{600} \cong 0.5$) and the other at late-stationary phase ($OD_{600} \cong 2.0$, about 17 hours after OD_{600} reached 0.5). Cells were pelleted in a JA10 rotor (Beckman) at 4500 x g for 20 minutes, washed twice in chilled lysis buffer (50 mM $NaPO_4$, 50 mM NaCl, pH 7.2) , and resuspended in lysis buffer containing 5 mM PMSF, 5 mM DTT and 5 mM p-aminobenzamidine.

Resuspended cells were lysed by two passes through a chilled large French press cell at 16,000 psi. The lysate was centrifuged at 17,000 x g for 30 minutes in a JA-20 rotor to remove the debris and intact cells. The supernatant was passed through a 0.45 μm polyethersulfone syringe filter (Whatman Puradisc^{TM}) prior to heparin-HPLC.

Heparin-HPLC

5 to 7 ml of the clarified cell lysate, containing ~50 mg of protein as estimated by Bradford assay, was loaded onto a 5 ml HiTrap Heparin HP column (Amersham Biosciences) on an AKTA Purifier system (Amersham Biosciences). The column was washed with 40 ml (8 column volumes) of 50 mM sodium phosphate buffer with 50 mM NaCl, pH 7.2, at which point the OD_{280} had returned to the baseline level. Heparin-bound proteins were eluted with a 60 ml (12 column volumes) 50 mM to 1M NaCl gradient in the same buffer and collected as twenty 3 ml fractions. Significant protein was only observed in the first 16 of these fractions by Bradford assay.

One-dimensional SDS polyacrylamide gel electrophoresis

One-third of every 3 ml elution fraction, which contained on average 80 µg of protein, was precipitated by addition of 500 µl of 50% TCA (final concentration of TCA is about 14%), washed twice by cold acetone, and resuspended in 20 µl SDS-PAGE sample loading buffer. Proteins in each sample were separated on 5 cm x 8 cm 10% Laemmli SDS polyacrylamide gels (37.5:1 acrylamide: bis-acrylamide) run at 20V/cm. The gels were stained with coomassie blue R-250 to visualize protein bands.

CyDye DIGE two-dimensional SDS polyacrylamide gel electrophoresis

CyDye DIGE two-dimensional PAGE was performed by the Protein Chemistry Laboratory at Texas A & M University. Heparin-binding proteins from either the exponential growth or stationary phase sample were pooled by mixing 200 µl from 16 protein-containing fractions. The pooled proteins were precipitated by 14% TCA,

washed twice with cold acetone and dried. Same amount of protein from exponential growth and stationary phase were combined as internal standard. Around 50 ng protein samples from exponential growth and stationary phase and the internal standard are each labeled with a different CyDye DIGE Fluor minimal dye. These labeled samples were then focused on Igphor immobilized gradient gels (14 cm pH 3–10 Nonlinear, Pharmacia) for 60–80,000 Volt hours, and separated by 12% SDS gels (14 x 14 cm) in the second dimension. This CyDye DIGE two-dimensional PAGE was repeated three times for statistic analysis. DeCyder™ v6.5 differential analysis software was used to visualize and analyze CyDye DIGE two-dimensional PAGE.

Montage in-gel digestion

459 coomassie-stained bands of approximately 2mm thick slices were excised from the Laemmli PAGE, 232 bands for the exponential growth sample, 227 bands for the stationary phase sample. In-gel digestions were carried out using the Montage In-Gel Digest Kit (Millipore).

In-solution sample digestion

200 µl aliquots of each heparin fraction were dialyzed three times for 4 hours each against 25 mM ammonium bicarbonate and subsequently dried on a speed vacuum at medium temperature. These were then reconstituted in 30 µl 50 mM ammonium bicarbonate, mixed with 2 µl of 20 µg/ml trypsin, and incubated overnight at 37°C. The samples were then frozen at -20°C prior to analysis.

Liquid chromatography

In-solution digested samples were analyzed using LC peptide separation and MALDI.

Mass spectrometry

Data was acquired in batch mode on the Applied Biosystems 4700. MS spectra were collected at 50 shots/spectrum, 40 sub-spectra per spot for a total of 2000 shots per spot using a random laser search pattern. The mass range was 800-5000 m/z with a focus mass of 2000 m/z. 15 MS/MS spectra per spot were collected at 70 shots/spectrum, 80 sub-spectra for a total of 5600 shots per spot. The mass window for MS/MS spectra was 150 relative resolution of the parent peak. The metastable ion suppressor and the collision cell were on with atmospheric gas present. LC spots were internally calibrated using the angiotensin I Mr 1296.6853 (Sigma) and glu-fibrinogen peptide B Mr 1570.6774 (Sigma). In-gel spots were externally calibrated with a mixture of Sigma angiotensin I Mr 1296.6853, ACTH 1-17 Mr 2093.0867, ACTH 18-39 Mr 2465.1989, and ACTH 7-38 Mr 3657.9294.

Analysis of spectral data was performed using the Applied Biosystems GPS Explorer Software Version 2.0. The parameters were as follows: taxonomy, *Escherichia coli*; database, Swiss Prot; enzyme, trypsin; maximum missed cleavages, 1; variable modifications, oxidation (Met), propionamide (Cys); peptide tolerance, 100 ppm; and a significance threshold of 90%. If systematic error is taken into account the ppm error for these externally calibrated spots is less than 40 ppm, however this was not feasible for

such a large sample set. Peak filtering for MS was from 800-5000 Da with a S/N limit of 15. Peak filtering for MS/MS was from 60 Da to 20 Da below each precursor mass with a minimum S/N filter of 8 and a mass tolerance of 0.3 Da. In-gel digestion spectra were searched using both MS and MS/MS data in "Combo mode" by Mascot (http://www.matrixscience.com from Matrix Science Ltd.). LC separation spectra were searched based on MS/MS data only by Mascot.

Positive in-gel identifications had a confidence interval ≥ 99.50 %, had a linear systematic error in the PMF assignments, and a lack of redundant PMF identifications (peptide masses matched to more than one identification). Positive LC MS/MS identifications had a protein confidence interval of ≥ 99.50 % and linear systematic error. Identifications based on single peptides were thrown out unless the protein was also found in other fractions of the same sample with more than 1 peptide.

Enhanced chemiluminescence (ECL) western blot

Proteins run on 10% Tris-Tricine gel were transfer to 0.2 μm nitrocellulose membrane using wire electrodes Trans-Blot cell (Bio-Rad) at 100 V for 3 to 4 hours. ECL Western blotting was performed according to manufacturer's protocol (SuperSignal West Pico, PIERCE). FtnA-specific polyclonal antisera were offered by Dr. Simon C. Andrews.

Validation of the results

Most of the proteins (70%) identified by in-gel/MS/MS have difference between ± 30% between the observed molecular weight (MW) and the theoretical molecular weight

(MW). All proteins identified by LC/MS/MS have 2 or more peptides. pI distribution of each fraction was also plotted to validate the chromatography behaviors of heparin-binding proteins on heparin columns. The peaks of the protein's pI in early fractions are 5-6, while those of late fractions are 10-11.

Bioinformatics

Data was organized and analyzed in a MySQL database, using Perl and PHP to provide a web-based user interface. This database is searchable and viewable at http://eep.tamu.edu/heparome. Identified proteins can be searched by gene name, SwissProt ID, b number, growth phases, or elution fraction number.

CHAPTER III

PROTEOMIC STUDIES BY AIX COLUMN AND 2D PAGE

Background

In *E. coli*, gene expression is tightly related to physiology. In response to various conditions, such as low pH, nutrient starvation, phage infection, antibiotics, etc., *E. coli* regulates gene expression in order to adapt to the changes (Arnold *et al.*, 2001; Hua *et al.*, 2004; Schembri *et al.*, 2003). Proteomics has been used in many areas to understand how protein expression patterns are related to the physiology of *E. coli* at the protein level, such as when cells enter stationary phase.

To increase the accuracy of protein identification, sample complexity should be reduced first. To separate complex protein mixtures and identify the proteins that present, non-denaturing liquid chromatography (NonDeLC) has been used (Badock *et al.*, 2001; Champion *et al.*, 2003; Ferguson and Smith, 2003). Under different conditions, a protein's chromatographic behavior will change if there are any covalent or non-covalent modifications, such as changes in binding partners. Coupled with mass spectrometry, NonDeLC can study proteins' chromatographic behavior and protein complex without denaturing the proteins. However, to study abundance and post-translational modifications, 2D PAGE (Henzel *et al.*, 1993; Langen *et al.*, 2000; Loo *et al.*, 2001; Tonella *et al.*, 2001) is an easier and more direct way to use. Two-dimensional gel electrophoresis (O'Farrell, 1975) is an traditional ways to separate complex protein mixtures containing thousands of components (Henzel *et al.*, 1993; Langen *et al.*, 2000;

Loo *et al.*, 2001; Tonella *et al.*, 2001).

In this study, these two methods were combined to study the soluble *E. coli* proteome from both exponential and stationary phase samples[1]. To study the differences in the soluble proteome from these two growth phases, many characters were investigated, such as protein patterns on 2D PAGE for fractions from anion exchange (AIX) chromatography, the elution profile for every protein from AIX column, abundance of proteins and proteins' chromatographic behavior.

Results

The strategy to study the *E. coli* proteome by combining NonDeLC, 2D PAGE, and mass spectrometry is shown in Fig. 3.1. The soluble *E. coli* proteins from both exponential growth and stationary phase were fractionated by AIX chromatography. An aliquot of each protein-containing fraction was separated by 2D PAGE. In total there were 19 2D gels run for each growth phase. Gel images were acquired and analyzed using the Z3 2D gel analysis system as described in Materials and Methods (Fig. 3.2). The amount of protein present in each spot was determined by comparison with horse myoglobin, which was added to each fraction as an internal marker of protein abundance.

---

[1]: In this chapter, protein samples were prepared by me, two-dimensional PAGE was performed by the Protein Chemistry Laboratory at Texas A & M University, and proteins on 2D PAGE were excised by me and identified by Dr. Sam Perkins from Dr. David Russell's laboratory using mass spectrometry.

Spots detected on only one gel and containing ≤ 100 ng of protein were filtered out. The observed molecular weight (Obs. MW) and observed pI (Obs. pI) of every spot were calculated using markers for molecular weight and standard curves from Pharmacia for pI. The numbers of spots visualized on individual 2D gels varied from 31 to 335 (Table 3.1). The 2D gels of early fractions had more protein spots than the gels of later fractions. Later fractions contained more acidic proteins (lower pI) (Fig. 3.3). These trends were the same for both growth phases.

To study abundance and elution pattern of proteins across a series of gels from one growth phase and then to compare different growth phases, protein spots on the 38 2D gels need to be connected, and spots on different gels corresponding to the same protein need to be inferred. This was done as described in Materials and Methods.

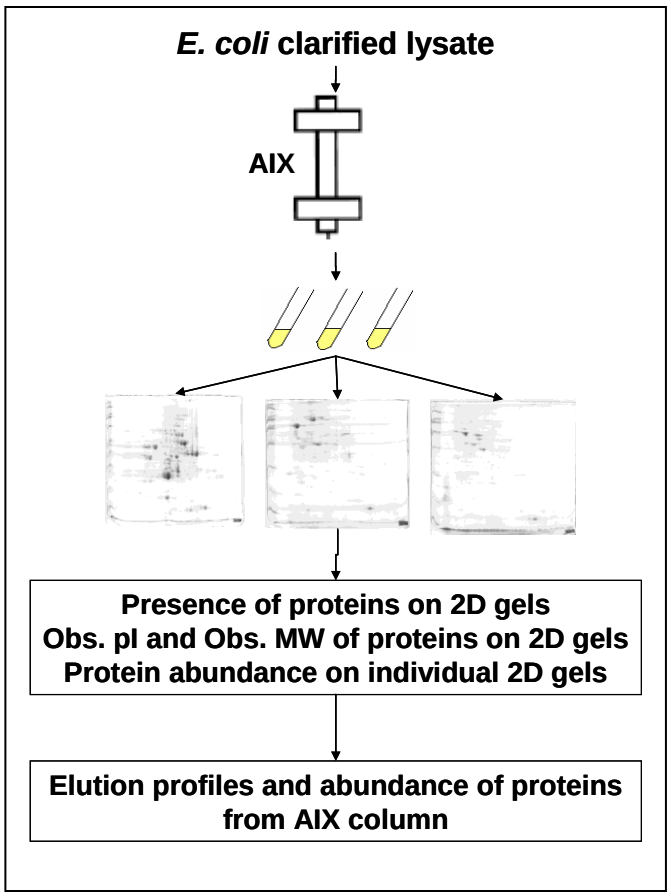
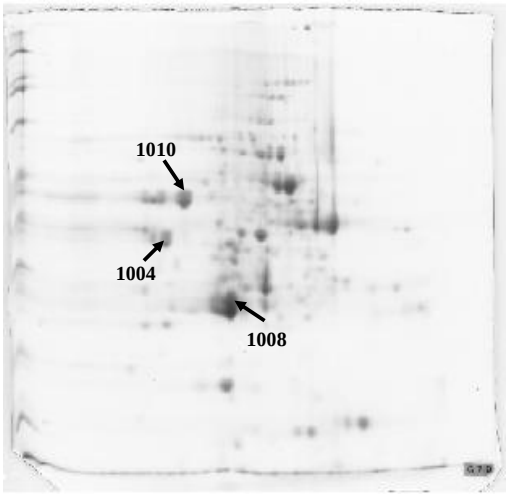


Figure 3.1. Flowchart of the strategy used to identify the soluble proteome of *E. coli* by LC and 2D PAGE. Clarified cell lysates were loaded onto mono Q HPLC columns. Bound proteins were eluted into 25 5 ml fractions. An aliquot of each fraction was separated on 2D PAGE. Spots on gels were analyzed using Z3 2D gel analysis systems combing with computer program. The identities of protein spots were determined by excising spots, in-gel digestion with trypsin, and MS/MS analysis.



ID	X	Y	Q
1008	886	839	41327
1010	764	602	12856
1014	707	672	6150

ID: Spot ID
X, Y: the coordinates of the spot in the gel image (in pixels)
Q: spot quantity

Figure 3.2. Analysis of 2D gels using Z3 2D gel analysis system. To each visualized spot, Z3 software assigns an ID number, locates the coordinates (X and Y), and calculates the spot intensity (Q).

Table 3.1. Number of spots visualized on every 2D PAGE

Exp.[a] Fraction #	Number of visualized spots	Stat.[b] Fraction #	Number of visualized spots
E5	249	S5	212
E6	243	S6	335
E7	205	S7	190
E8	216	S8	236
E9	227	S9	167
E10	145	S10	183
E11	204	S11	232
E12	213	S12	200
E13	231	S13	172
E14	261	S14	223
E15	149	S15	155
E16	97	S16	101
E17	107	S17	127
E18	111	S18	124
E19	61	S19	56
E20	64	S20	62
E21	47	S21	64
E22	94	S22	31
E23	125	S23	121

a: exponential phase
b: stationary phase

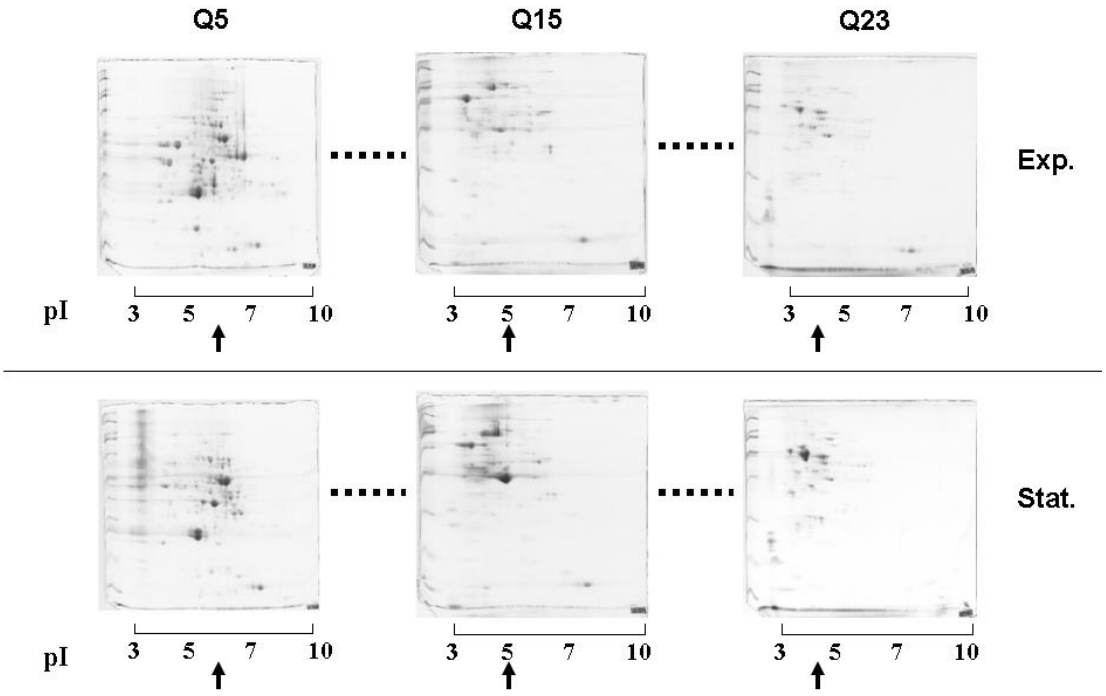


Figure 3.3. 2D PAGE of fractions from AIX HPLC for both exponential growth and stationary phase. Numbers labeled below gels are pI range and arrows are the pI of the majority of spots on gels.

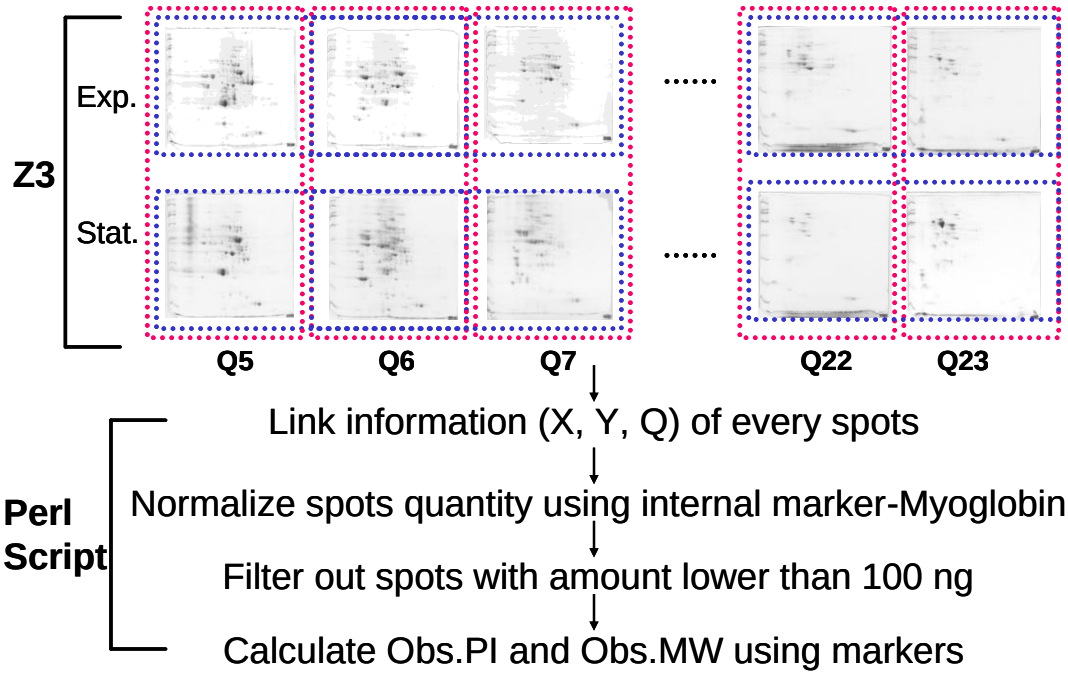


Figure 3.4. Strategy to analyze a series of 2D PAGE. 2D gels from both exponential and stationary phases were connected using the Z3 software, which assigns an ID number, locates the coordinates (X and Y), and calculates the spot intensity (Q). Perl scripts were applied to calculate spots quantity, observed molecular weight and observed pI.

In total, elution profiles for 1949 nonredundant spots were identified. 734 of these were more abundant in exponential growth, 920 were more abundant in stationary phase and 295 were present in equivalent amount in both growth phases. Details of visualized spots, such as elution profile, Obs. MW, Obs. pI and amount, can be found in EEP (http://eep.tamu.edu/2DPage/index.php?page=results2.html) (Fig. 3.5 and Fig. 3.6). After comparing the abundance of spots between these two growth phases, spots with big difference of amount (ratio >10 or only in one growth phase with amount > 2 μg) were chose to be analyzed further. In total 193 selected spots were cut and identified by MS/MS. Some spots did not have protein identified, while some spots were identified as same proteins. Overall, 102 cut spots have proteins identified (Table 3.2A), and 91 cut spots have no protein identified (Table 3.2B). In total 61 proteins were identified (Table 3.3). Some proteins were found only in one protein spot, such as DnaK (Fig. 3.7), other proteins were found in multiple protein spots, such as AceA (Fig. 3.8). AceA were found in two protein spots with similar elution positions (Fig. 3.8A) but different abundance trends between exponential growth and stationary phase (Fig. 3.8B). In total, 34 of the protein identified were more abundant in exponential growth, 25 were more abundant in stationary phase, and 2 have equivalent amount in both growth phases (Table 3.4).

Spots #	Obs. PI	Obs.MW (kD)	Ratio (L/S)	Cell	Total (ug)	Fraction #																
						5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
3085	4.9	24.3	2.01	L	57.1	0	0	0	0	0	0	1.7	8.7	38	8.5	0.4	0	0	0	0	0	0	0
				S	28.4	0	0	0	0	0	1.8	2.5	18	5.7	0.6	0	0	0	0	0	0	0	0
1647 a	6	58.1	2.02	L	4.62	0	0.3	1.9	1.8	0.6	0.1	0	0	0	0	0	0	0	0	0	0	0	0
				S	2.29	1.9	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2358	5.5	42.75	2.02	L	1.19	0	0.4	0	0.6	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0
				S	0.59	0	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2_1141	5.3	45.05	2.03	L	5.97	0	0	0	0	0	0	0	0	0.6	2	2.1	1.2	0.1	0	0	0	0	0
				S	2.94	0	0	0	0	0	0	0	0	1.3	1.7	0	0	0	0	0	0	0	0

Figure 3.5. Example of elution profiles for spots visualize on a series of 2D PAGE. Spot ID, Obs. pI, Obs. MW, amount ratio between exponential growth and stationary phase, growth phase, total amount, and spots amount (µg) in every fractions from AIX column were displayed.

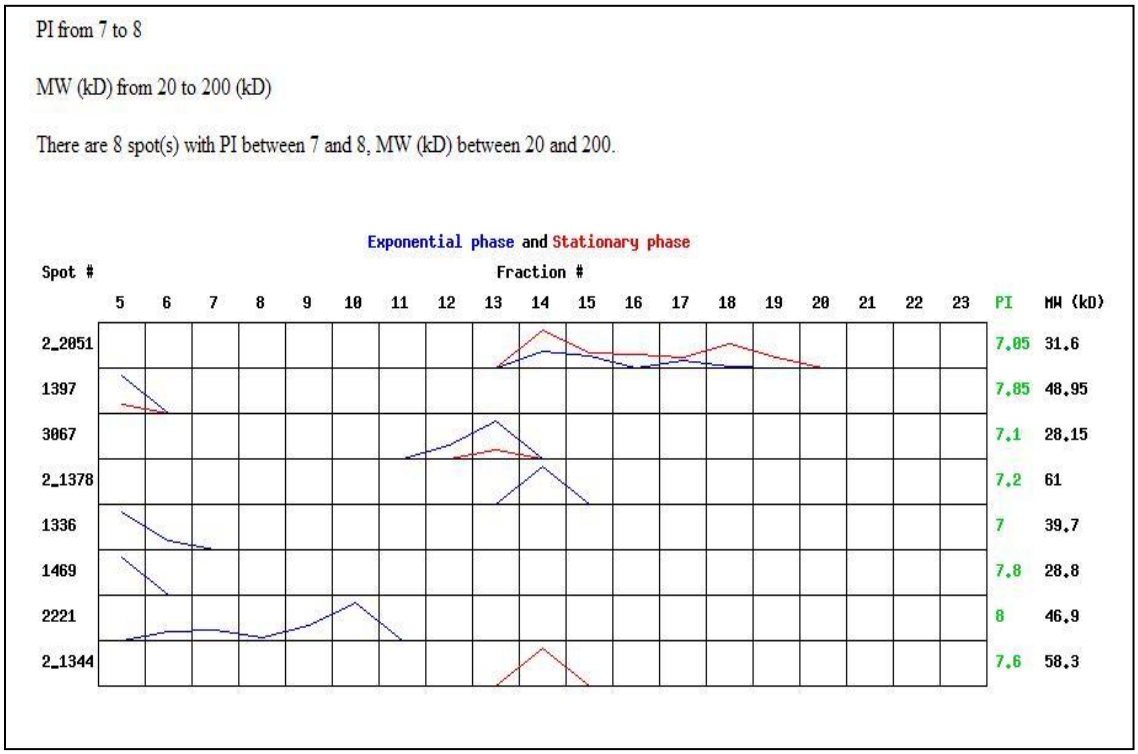


Figure 3.6. Elution profiles of proteins with pI from 7 to 8 and with MW from 20 to 200 kD. Proteins from eight spots on 2D PAGE were found. Blue line refers sample from exponential growth and red line refer stationary phase. Obs. MW and Obs. pI were on the right side of elution profile for every spot.

Table 3.2A. Cut spots with protein identified.

Spot_ID	Gene	b number	Spot_ID	Gene	b number	Spot_ID	Gene	b number
2_2295	aceA	b4015	2_1658	fusA	b3340	1585	purC	b2476
2_1911	aceA	b4015	2_1501	fusA	b3340	1754	pykF	b1676
1499	aceE	b0114	2458	galM	b0756	1689	ribB	b3041
1413	acnB	b0118	3_1686	gdhA	b1761	2_1770	rpsA	b0911
1416 a	acnB	b0118	2651	ghrB	b3553	2_1148 a	rpsA	b0911
1392	acnB	b0118	1678	glnS	b0680	2_1450	rpsA	b0911
1470	acnB	b0118	1686	glnS	b0680	2_1148 b	rpsA	b0911
1976	adhE	b1241	2177	gltD	b3213	1052 a	serA	b2913
3079	ahpC	b0605	2341	glyA	b2551	2256	serA	b2913
3053	ahpC	b0605	2254	gnd	b2029	2296	serA	b2913
2398	aspC	b0928	2307	gnd	b2029	1172 e	serC	b0907
1743	aspS	b1866	4_1183	groL	b4143	1479	sucA	b0726
1363 a	atpA	b3734	2_2165	groS	b4142	2_1814	tig	b0436
2_1823	atpA	b3734	2034	guaB	b2508	2_2166 a	tig	b0436
2_1327 a	atpA	b3734	2028	leuA	b0074	2_1399	tig	b0436
1426	atpA	b3734	1299	leuD	b0071	2_1341	tig	b0436
1648	atpG	b3733	3086	luxS	b2687	1218	tktA	b2935
3150	bcp	b2480	3083	mdaB	b3028	1434	tpx	b1324
2792	cbpA	b1000	1843	mdoG	b1048	2779	tsf	b0170
3115	crr	b2417	1567	metE	b3829	1545	tsf	b0170
1089	cysK	b2414	2900	metF	b3941	1142	tsf	b0170
2758	cysK	b2414	2974	mtn	b0159	1096	tsf	b0170
2859	dapD	b0166	1048	nadE	b1740	4_2227	tufA	b3339
2_1735	dnaK	b0014	3165	ndk	b2518	2_1272	tufA	b3339
2_1948	dnaK	b0014	1609	pflB	b0903	2_1940	tufA	b3339
1185	dnaK	b0014	2613	pflB	b0903	2471	tufA	b3339
3133	dps	b0812	1222	pgk	b2926	2_1232	tufA	b3339
3146	efp	b4147	3_1827	pnp	b3164	2_1959	tufA	b3339
2516	eno	b2779	2720	potD	b1123	2_1963	tufA	b3339
2_1593	fusA	b3340	3066	ppa	b4226	2_2177 a	uspA	b3495
2_1474	fusA	b3340	1151	proS	b0194	1339	yhgF	b3407
1574	fusA	b3340	1117	proS	b0194	1772	ynjA	b1753
2_1566	fusA	b3340	3117	proS	b0194	2_1139	ynjA	b1753
2_1589	fusA	b3340	1650	purC	b2476	3009	yodA	b1973

Table 3.2B. Cut spots without protein identified.

Spot_ID	Spot_ID	Spot_ID	Spot_ID	Spot_ID	Spot_ID
1066	1430	2124	2743	2_1161	2_1842
1070	1441	2164	2769	2_1220	2_1868
1078	1550	2246	2805	2_1224	2_1989
1085	1553	2269	2810	2_1266	2_1995
1103	1558	2305	2819	2_1325	2_2104
1170	1628	2323	2863	2_1405	2_2471 a
1171	1673	2338	2866	2_1419	3_1621
1186	1706	2425	2873	2_1541	3_1763
1263	1749	2436	2890	2_1562	3_1765
1289	1826	2456	2892	2_1573	3_1776
1298	1955	2484	3082	2_1701	4_1339
1300	1998	2507	3170	2_1751	4_1717
1326	2047	2535	1628 b	2_1774	4_2181
1346	2103	2554	2_1075	2_1792 a	4_2321
1364	2104	2732	2_1137	2_1792 b	4_2329

Table 3.3. Proteins identified from 2D PAGE.

Gene Name	b Number	SwissProt	Cal. pI[a]	Obs. pI[b]	Cal. MW (kD)[c]	Obs. MW (kD)[d]	Total Amount in Exp. [e] (µg)	Total Amount in Stat. [f] (µg)
aceA	b4015	P05313	5.16	4.90; 5.10	47.52	49.55; 45.60	3.64	1.49
aceE	b0114	P06958	5.46	5.4	99.54	256.05	13.91	1.22
acnB	b0118	P36683	5.24	5.20; 5.30; 5.45; 5.50	93.5	103.90;103.65; 110.75; 107.30	0.46	7.08
adhE	b1241	P17547	6.33	6.1	96	101.8	2.24	0.12
ahpC	b0605	P26427	5.03	4.9	20.63	24.6	2.29	24.82
aspC	b0928	P00509	5.54	5.6	43.57	42.6	0.1	1.04
aspS	b1866	P21889	5.47	5.6	65.91	69.4	0.03	0.54
atpA	b3734	P00822	5.8	5.80; 5.90; 6.00	55.22	79.90; 79.10; 50.95; 40.00	20.05	5.39
atpG	b3733	P00837	8.84	8.35	31.58	29.05	1.14	0.06
bcp	b2480	P23480	5.03	4.95	17.63	16.6	0.57	0.03
cbpA	b1000	P36659	6.33	6.4	34.46	33.9	0.02	0.75
crr	b2417	P08837	4.73	4.5	18.12	20.7	0	16.43
cysK	b2414	P11096	5.83	5.70; 5.95	34.36	35.40; 34.50	22.36	0.2
dapD	b0166	P03948	5.56	5.45	29.89	32.75	4.76	0.26
dnaK	b0014	P04475	4.83	4.70; 4.75; 5.90	68.98	78.00; 76.90; 43.80	0.22	12.2
dps	b0812	P27430	5.72	5.6	18.56	17.45	0.17	2.04
efp	b4147	P33398	4.9	4.3	20.46	16.9	2.64	0
eno	b2779	P08324	5.32	5.1	45.52	43.5	2.58	0

Table 3.3. Continued

Gene Name	b Number	SwissProt	Cal. pI[a]	Obs. pI[b]	Cal. MW (kD)[c]	Obs. MW (kD)[d]	Total Amount in Exp. [e] (µg)	Total Amount in Stat. [f] (µg)
fusA	b3340	P02996	5.24	5.00; 5.05; 5.10; 5.40	77.45	101.75; 93.95; 89.30	16.59	23.64
galM	b0756	P40681	4.84	4.8	38.19	40.35	0.09	2.69
gdhA	b1761	P00370	5.98	6.05	48.58	44.3	0.06	2.79
ghrB	b3553	P37666	5.5	5.45	35.4	38.3	0.97	0.03
glnS	b0680	P00962	5.89	5.80; 5.90	63.35	68.85; 67.95	3.74	0.21
gltD	b3213	P09832	5.54	5.6	51.88	52.3	0	4.41
glyA	b2551	P00477	6.03	6.1	45.32	43.4	2.04	0
gnd	b2029	P00350	5.05	5.1	51.48	45.6	0.92	15.7
groL	b4143	P06139	4.85	4.9	57.2	48.3	2.77	29.34
groS	b4142	P05380	5.15	4.85	10.39	15.3	0.03	2.05
guaB	b2508	P06981	6.02	5.8	52.02	55.15	2.39	0.69
leuA	b0074	P09151	5.47	5.4	57.17	55.75	0.77	0.04
leuD	b0071	P30126	5.16	5.3	22.36	24.3	2.48	0
luxS	b2687	P45578	5.18	5.25	19.29	23.05	0.54	6.53
mdaB	b3028	P40717	5.85	5.9	21.89	24.3	0	4.54
mdoG	b1048	P33136	6.26	6.3	55.37	59.45	1.64	0.13
metE	b3829	P25665	5.61	5.7	84.54	119.3	8.79	0.72
metF	b3941	P00394	6	6	33.1	32	0	2.23
mtn	b0159	P24247	5.09	5.2	24.35	28.1	0.19	3.75
nadE	b1740	P18843	5.41	5.6	30.64	36.7	5.73	0.23
ndk	b2518	P24233	5.55	5.6	15.33	14	3	0
pflB	b0903	P09373	5.69	5.70; 5.75	85.23	84.00; 37.75	4.95	0.31
pgk	b2926	P11665	5.08	5.8	40.99	82.1	0	2.66
pnp	b3164	P05055	5.11	5.2	77.1	50.6	2.03	0
potD	b1123	P23861	4.86	4.5	36.49	36.3	7.61	0
ppa	b4226	P17288	5.03	5	19.57	24.4	0	18.29

Table 3.3. Continued

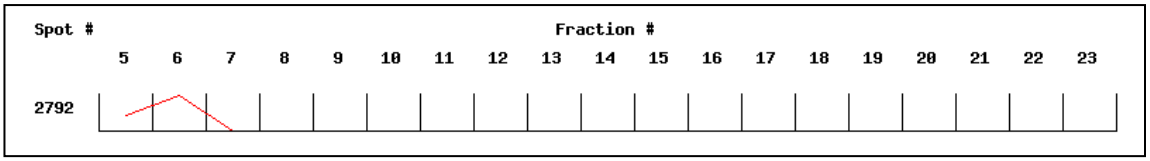
Gene Name	b Number	SwissProt	Cal. pI[a]	Obs. pI[b]	Cal. MW (kD)[c]	Obs. MW (kD)[d]	Total Amount in Exp. [e] (µg)	Total Amount in Stat. [f] (µg)
proS	b0194	P16659	5.12	4.60; 5.20	63.69	72.15; 68.10; 21.50	0.59	15.23
purC	b2476	P21155	5.07	5.1	27	29.25; 28.8	3.37	0.1
pykF	b1676	P14178	5.77	5.8	50.73	57.5	2.7	0
ribB	b3041	P24199	4.9	4.95	23.35	29.35	0.53	18.8
rpsA	b0911	P02349	4.89	4.40; 4.60; 4.70	61.16	78.15; 74.10; 70.10; 69.50	85.79	7
serA	b2913	P08328	5.93	5.90; 5.95	44.04	46.20; 42.80	12.59	0.62
serC	b0907	P23721	5.37	5.5	39.65	40.8	2.62	0
sucA	b0726	P07015	6.04	6	105.06	110.45	0.85	0.05
tig	b0436	P22257	4.83	4.30; 4.40; 4.50; 4.60	48.19	63.10; 58.20; 53.95; 52.50	4.1	2.61
tktA	b2935	P27302	5.43	5.55	72.21	77.45	0.19	4.8
tpx	b1324	P37901	4.75	4.9	17.7	19.9	4.79	0
tsf	b0170	P02997	5.22	5.20; 5.30	30.29	35.00; 34.20; 33.80	16.41	0.09

Table 3.3. Continued

Gene Name	b Number	SwissProt	Cal. pI[a]	Obs. pI[b]	Cal. MW (kD)[c]	Obs. MW (kD)[d]	Total Amount in Exp. [e] (µg)	Total Amount in Stat. [f] (µg)
tufA	b3339	P02990	5.3	4.10; 5.20; 5.30; 5.40; 5.45	43.18	49.65; 47.70; 45.20; 43.65; 43.30; 43.10; 36.30	5.13	58.82
yhgF	b3407	P46837	5.92	5.95	85.12	88.35	2.1	0.06
ynjA	b1753	P76222	8.8	5.40; 5.70	20.53	75.35; 44.60	3.14	0.26
yodA	b1973	P76344	5.66	5.6	22.34	28.5	0	3.54

a: Cal. pI: calculated pI.
b: Obs. pI: observed pI.
c: Cal. MW: calculated molecular weight.
d: Obs. MW: observed molecular weight.
e: Exp: exponential growth.
f: Stat: stationary phase.

A)

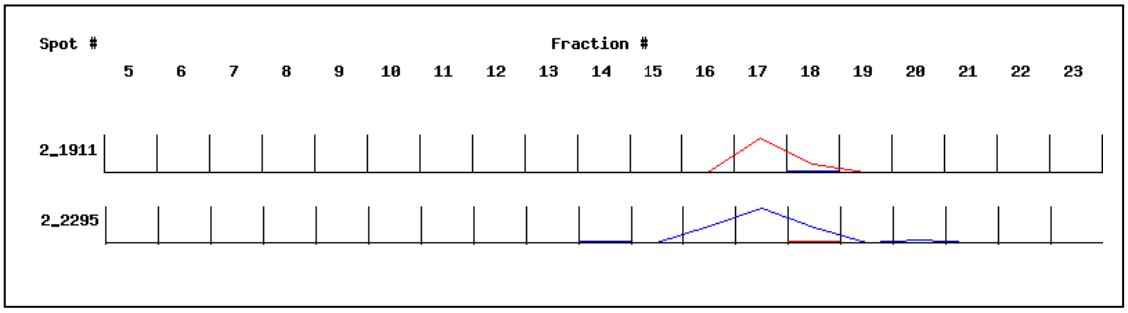


B)

Spot ID	Obs. PI	Obs. MW (kD)	Log2(E/S)	Exp. Total (µg)	Stat. Total (µg)
2792	6.4	33.9	-5	0.02	0.75
Total			**-5**	**0.02**	**0.75**

Figure 3.7. Information of the protein spot identified as DnaK on 2D PAGE. A) Elution profile of the protein spot (spot ID 2792) identified as DnaK. B) Protein abundance trend of spot 2792 between exponential and stationary phase. Protein amount in protein spot 2792 was found more in stationary phase.

A)



B)

Spot ID	Obs. PI	Obs. MW (kD)	Log2(E/S)	Exp. Total (µg)	Stat. Total (µg)
2_1911	4.9	45.6	-4	0.08	1.37
2_2295	5.1	49.55	+5	3.56	0.12
Total			**+1**	**3.64**	**1.49**

Figure 3.8. Information of protein spots identified as AceA on 2D PAGE. A) Elution profiles of protein spots (spot ID 2_1911 and 2_2295) identified as AceA. B) Protein abundance trends of spot 2_1911 and 2_2295 between exponential and stationary phase. Protein amount in protein spot 2_1911 was found more in stationary phase. Protein amount in protein spot 2_2295 was found more in stationary phase.

Table 3.4. The summary of proteins identified from 2D PAGE.

More amount in Exp.	One spot	*aceE, adhE, atpG, bcp, dapD, efp, eno, ghrB, glyA, guaB, leuA, leuD, mdoG, metE, nadE, ndk, pnp, potD, pykF, serC, sucA, tpx, uspA, yhgF*
	Multiple spots with same trend	*atpA, cysK, glnS, pflB, purC, serA, tsf, ynjA*
	Multiple spots with different trend	*aceA, rpsA*
More amount in Stat.	One spot	*aspC, aspS, cbpA, crr, dps, galM, gdhA, gltD, groL, groS, luxS, mdaB, metF, mtn, pgk, ppa, ribB, tktA, yodA*
	Multiple spots with same trend	*acnB, dnaK, gnd, proS*
	Multiple spots with different trend	*ahpC, tufA*
Similar amount in both	Multiple spots with different trend	*fusA, tig*

Previously, our laboratories had analyzed the *E. coli* exponential phase proteome using two-dimensional non-denaturing liquid chromatography (2D-NonDeLC) and mass spectrometry (Champion *et al.*, 2003). 2D-NonDeLC applied anion exchange (AIX) column and hydrophobic interaction column (HIC) to separate protein mixture. Since the first dimension of NonDeLC was the same as this study used, some properties of identified proteins are expected to be same as those from LC/2D PAGE. 45 of the 61 proteins identified by this study were also detected in the earlier study (unpublished NonDeLC study by Matthew Champion). For these proteins we could compare the elution profiles directly. For 12 proteins the elution profiles were very similar (Fig. 3.9).

To compare proteins in every fraction from NonDeLC and LC/2D PAGE, virtual 2D PAGE were generated using the theoretical pI and molecular weight information of proteins identified by 2D NonDeLC (http://eep.tamu.edu/nondelc/results/). The real 2D gels have more protein spots and more complicated patterns (compare the gels in Fig. 3.10A and 3.10B). The number of proteins could be identified on 2D PAGE is bigger then that by 2D NonDeLC. Furthermore, to make the virtual 2D PAGE for NonDeLC and real 2D gel more comparable, every real 2D PAGE was redrawn with spots labeled with their identified proteins using the observed pI and observed molecular weight of every spot. By comparing the virtual 2D PAGE and redrawn real 2D PAGE (Fig. 3.10C) for same fractions, some proteins were identified by both methods, however each methods has their own unique protein identification. Number of proteins identified on 2D PAGE is bigger then that by 2D NonDeLC.

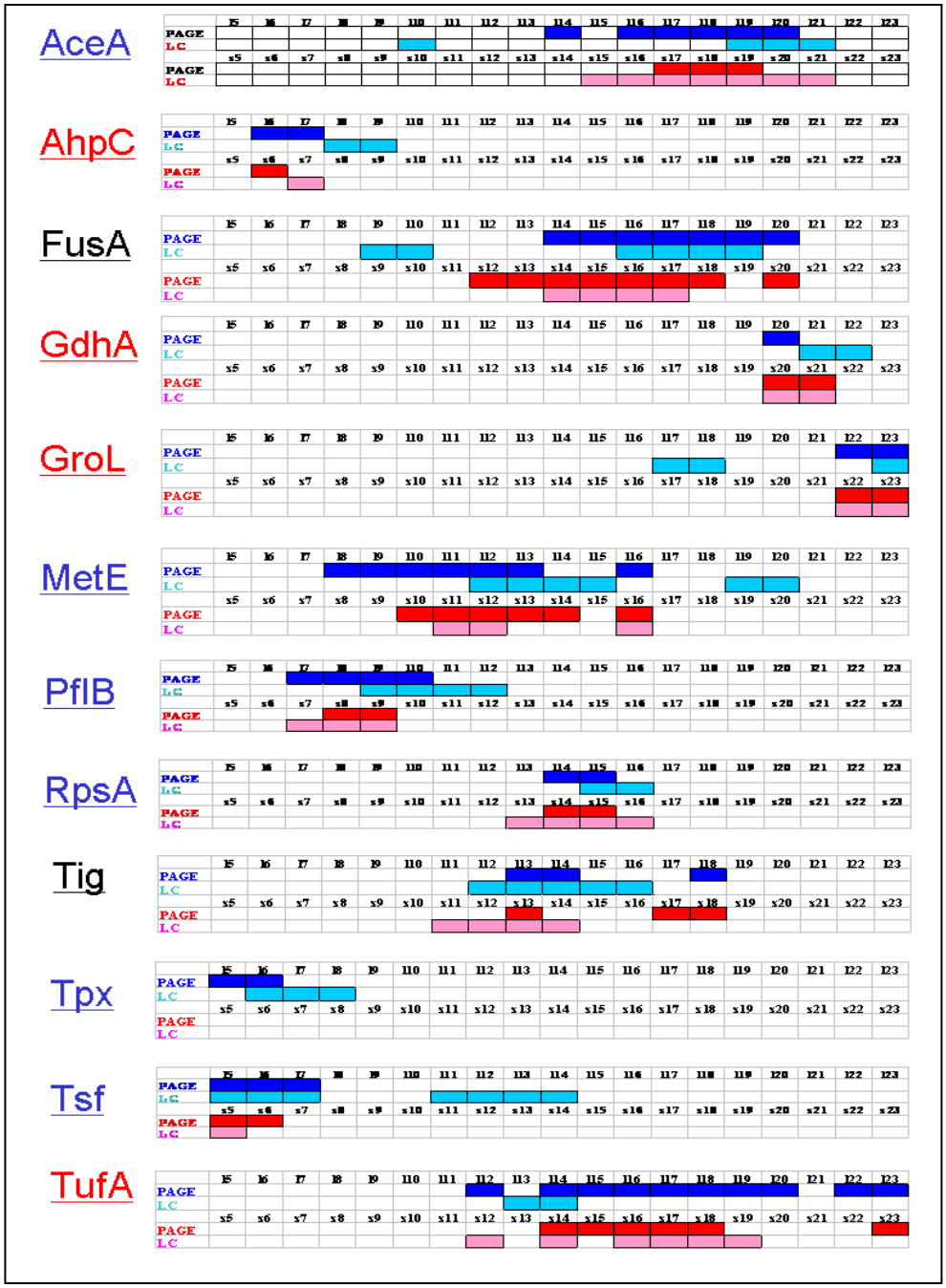
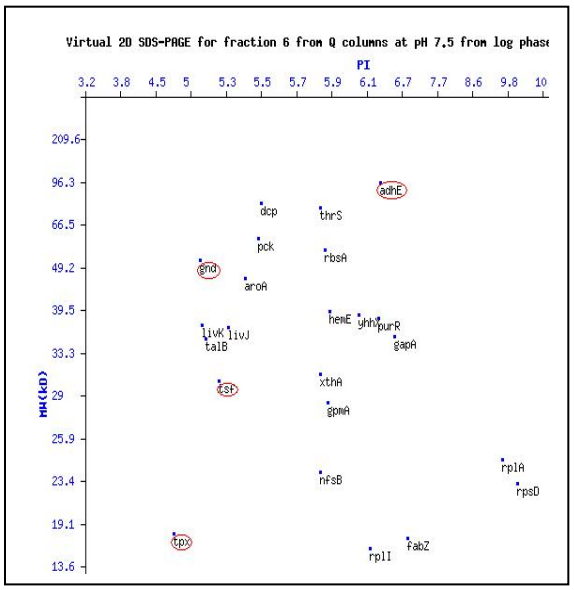
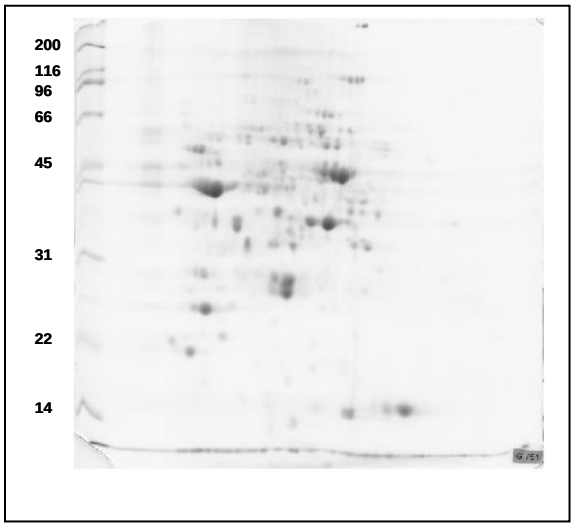


Figure 3.9. Comparison of elution profiles between 2D PAGE and 2DLC. Existences of protein in certain fraction in exponential growth are boxes filled with blue (by 2D PAGE) and light blue (by 2D LC). Stationary phase are boxes filled with red (by 2D PAGE) and pink (by 2D LC).

A)



B)



C)

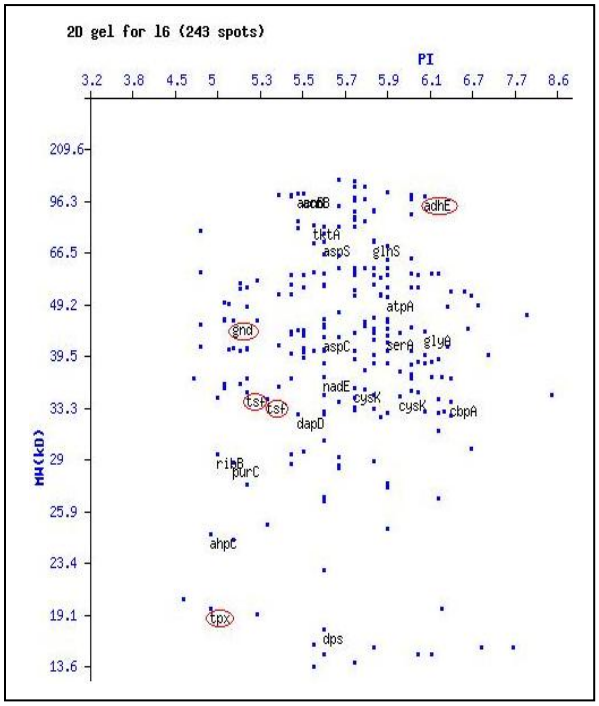


Figure 3.10. The virtual 2D PAGE (A), the real 2D PAGE (B) and the redrawn real 2D PAGE (C) for 6th elution fraction from AIX column for exponential growth sample of *E. coli*. 23 protein spots were visualized on the virtual 2D PAGE. 243 protein spots were visualized on the real and redrawn 2D PAGE. Proteins labeled on C were identified by in-gel digestion followed with tandem mass spectrometry. Proteins circled were identified by both methods.

Discussion

To study the differences in the soluble proteome from both exponential and stationary phase samples, NonDeLC and 2D PAGE were applied. Proteins in 19 fractions from AIX chromatography were separated on 2D PAGE for sample from each growth phase. Using Z3 2D gel analysis system and computational languages, thousands of spots on 38 2D PAGE were visualized and analyzed. Of a total of 1949 nonredundant spots, 38% of them were more abundant in exponential growth, 47% of them were more abundant in stationary.

For the 179 protein spots cut from gels and identified by MS/MS analysis, 25 proteins (Table 3.5) were more abundant in stationary phase. The products of 4 genes (*cbpA, dnaK, groL, and groS*) with more amounts in stationary phase are chaperone proteins. Chaperone proteins, whose function is to assist proteins to fold properly, are often involved in to different stresses such as heat shock (Hightower, 1991). DnaK, GroL and GroS proteins are heat shock protein in *E. coli* (Gragerov *et al.*, 1992; Tilly *et al.*, 1983). CbpA protein, a DnaJ homolog, is a DnaK co-chaperone responding to heat shock (Bird *et al.*, 2006; Chae *et al.*, 2004). Microarray analysis showed that transcription of cbpA gene is up-regulated under starvation (Tani *et al.*, 2002) and up-regulated by RpoS (Patten *et al.*, 2004). It has been known that *E. coli* cells increase the level of chaperone proteins to decrease protein miss-folding when cell face different stresses entering stationary phase (Yoon *et al.*, 2003; Zhang *et al.*, 2006).

The products of 5 genes (*crr, galM, gnd, pgk, tktA*) involved in carbohydrate metabolism are more abundant in stationary phase. Study showed that *crr*, which is

involved in phosphotransferase system (Roseman and Meadow, 1990), is up regulated by sigma S (Lacour and Landini, 2004). Gene *tktA* is one of the two genes encoding transketolase isoenzymes. Study has showed that at transcriptional level, *tktA* gene was suppressed in the stationary growth phase(Jung *et al.*, 2005). In our study, product of *tktA* was more abundant in stationary phase. This could be caused that TktA protein is more stable in stationary phase, which could be test by pulse-chase experiments.

The products of 3 genes (*ahpC, dps, and mdaB*) involved in starvation or stress response are more abundant in stationary phase. *dps* is known to be unregulated by RpoS (Dukan and Touati, 1996; Lacour and Landini, 2004; Patten *et al.*, 2004). Dps protein, as a nonspecific DNA binding protein and a bacterial ferritin, known with more abundant in stationary phase (Weichart *et al.*, 2003), is involved in resistance to oxidative stress by sequestering free ion and protecting DNA from damage. Microarray analysis also showed that transcription of *dps* gene is up-regulated under starvation (Tani *et al.*, 2002) and up-regulated by RpoS (Patten *et al.*, 2004). AhpC protein, as an alkyl hydroperoxide reductase, is regulated by OxyR. Some study showed that AhpC has similar amount between exponential growth and early stationary phase, but degrading in late stationary phase in *E. coli* (Weichart *et al.*, 2003). We found that the amount of AhpC protein was increased after 4 hours entering stationary phase. Since AhpC protects cells from damage caused by hydroperoxide (Charoenlap *et al.*, 2005; Jung and Kim, 2003) in and reactive nitrogen intermediates (Chen *et al.*, 1998), it is possible that *E. coli* cells increase level of AhpC protein in early stationary phase to protect themselves and degraded in late-stationary phase after other mechanisms adapt to stationary phases. To test this, amount

of AhpC protein at different growth phases, such as exponential phase, early-stationary phase, and late stationary phase, could be detected by Western or PAGE coupled with mass spectrometry. MdaB (modulator of drug activity B) protein is a NADPH quinone reductase, which has resistance to the antibiotics DMP 840, adriamycin and etoposide (Adams *et al.*, 2005). Recent study shows that structure of MdaB is similar to many flavoproteins (Adams and Jia, 2006). Induction of proteins Dps, AhpC, and MadB protects *E. coli* cells from oxidative and drug stress.

The products of three genes (*aspC, gdhA, gltD*) involved in amino acid metabolism are more abundant in stationary phase. It has been shown that GdhA protein, a glutamate dehydrogenase, is significantly up-regulated under nitrogen starvation of *E. coli* cells (Kabir *et al.*, 2004). Gene of *gltD*, a gene for glutamate synthase, was found induced by phosphate starvation in *E. coli* cells (Metcalf *et al.*, 1990). AspC protein, which is involved in amino acid degradation, might be also induced by nitrogen starvation.

Table 3.5. Proteins more abundant in stationary phase identified by 2D PAGE.

Gene	B_num[a]	Product
Amino acid metabolism		
aspC	b0928	aspartate aminotransferase, PLP-dependent
gdhA	b1761	glutamate dehydrogenase, NADP-specific
gltD	b3213	glutamate synthase, 4Fe-4S protein, small subunit
Chaperone		
cbpA	b1000	curved DNA-binding protein, DnaJ homologue that functions as a co-chaperone of DnaK
dnaK	b0014	chaperone Hsp70, co-chaperone with DnaJ
groL	b4143	GroEL, chaperone Hsp60, peptide-dependent ATPase, heat shock protein
groS	b4142	GroES, chaperone binds to Hsp60 in pres. Mg-ATP
Carbohydrate metabolism		
crr	b2417	phosphotransferase system, glucose-specific enzyme IIA
galM	b0756	galactose-1-epimerase (mutarotase)
gnd	b2029	gluconate-6-phosphate dehydrogenase, decarboxylating
pgk	b2926	phosphoglycerate kinase
tktA	b2935	glycolysis; gluconeogenesis
Starvation and stress response		
ahpC	b0605	alkyl hydroperoxide reductase, C22 subunit
dps	b0812	Fe-binding and storage protein; stationary phase nucleoid protein that sequesters iron and protects DNA from damage
Translation		
aspS	b1866	aspartyl-tRNA synthetase
proS	b1094	prolyl-tRNA synthetase
tufA	b3339	protein chain elongation factor EF-Tu (duplicate of tufB)
Miscellaneous function		
acnB	b0118	bifunctional aconitate hydratase 2 and 2-methylisocitrate dehydratase
luxS	b2687	autoinducer 2 (AI-2) synthase; acts in quorum sensing
metF	b3829	5,10-methylenetetrahydrofolate reductase
mtn	b0159	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase
ppa	b4226	inorganic pyrophosphatase
ribB	b3041	3,4-dihydroxy-2-butanone-4-phosphate synthase
yodA	b1973	cadmium-induced metal binding protein

a: Blattner number

The increased abundance of some proteins was surprising. AspS, ProS, and TufA are parts of translation machinery. AspS and ProS are two aminoacyl-tRNA synthetases. Usually, whit low growth-rate or nutrient limitation, aminoacyl-tRNA synthetases will be degraded or expressed less. One reason might because more AspS and ProS flow through the AIX column in exponential phase. To test if more AspS and ProS are in flow through from AIX column, PAGE couple with MS could be applied. Gene of *tufA* encodes a protein chain elongation factor, EF-Tu. TufA protein is often found up-regulated in exponential growth of *E. coli* cells, while in our study TufA was more abundant from AIX column for stationary phase. This could due to more TufA bound to AIX column in stationary phase. To test it, proteins in flow through fractions could be run on 2D PAGE and TufA could be identified by MS.

About 90% the observed molecular weights of proteins identified by this study on 2D PAGE are similar (±20%) to their theoretical molecular weights. The ones with observed molecular weight lower than their theoretical one could be due to protein degradation. Some proteins, such as AceE, AtpA, Pgk and YnjA, have observed molecular weights that are higher than the expected theoretical molecular weight. This aberrant migration might be caused by charges of amino acids in proteins (Armstrong and Roman, 1992). About 95% the observed pIs of identified proteins are similar (±10%) to the theoretical pI. Proteins, such as DnaK, TufA and YnjA, with higher or lower observed pI could have post-translational modification or were degraded. Modifications could include biologically significant post-translational modifications, but modifications during protein handling for 2D gels are common (Herbert *et al.*, 1998; Westermeier *et al.*,

1983). In this study, DnaK was found in 3 different protein spots. One of these 3 protein spots have higher observed pI, which might be caused by protein degradation since this spots had lower observed MW (43.8 kD) compared with the calculated one (68.98 kD). Observation of these characters for DnaK was also saw in other literature (Link *et al.*, 1997).

Based on the comparison between 2D NonDeLC and AIX/2D PAGE studies, elution profiles for same proteins are similar but not exactly same. 2D PAGE has more protein spots and more complicated patterns which indicate that more proteins would be identified after more efficient separation by 2D PAGE. This also could be post-translational modification and/or protein degradation. Although 2D PAGE study is labor intensive, it can separate complex protein mixture more efficiently. Moreover, more properties of protein can be studied by 2D PAGE such as protein amount, post-translational modifications, etc.

This is a promising pilot analysis of large number of proteins on a series of 2D gels. About 10% of all unique spots were cut and identified. Still spots with amount difference lower than 10 fold, and spots have similar amount between these two growth phases were not cut and identified by MS. One possible way to predict proteins of uncut spots is to compare patterns of protein distributed on 2D gels of this study with other 2D gels studies, such as SwissProt, CyberCell Project. This way could be a reference but will not be very accurate. One reason is because different 2D gels studies used different *E. coli* strains or different growth phases, which have different proteome. Another reason is the pI range of 2D gels used by theses studies are not exactly same, which might cause

different patterns of protein distributed. Also *E. coli* proteome have many proteins with similar pI and molecular weight, so it is very hard to differentiate proteins even in a very small area. However, if combined with more robotic gel cutting and MS identification, more or all proteins can be identified on 2D PAGE.

There are still problems need to be solved. One problem is that the success of protein identification using in-gel digestion coupled with MS/MS is relatively low. 91 out of 193 cut spots did not have protein identified. One possible reason is that the quality of protein identifications for these 91 spots is lower than the threshold (a confidence interval ≥ 99.50 %, linear systematic error in the PMF assignments, and a lack of redundant PMF identifications). For this reason, success of protein identification could be increased by adjusting the threshold. Another possible reason is the efficiency of in-gel digestion is low. There are also some other problems. For example, spots at same position wouldn't be same proteins all the times, because *E. coli* cells have proteins with same observed pI and observed molecular weight. Same proteins showing in distant fractions could be recognized as different proteins. Same protein sometime has multiple spots on same gel, which is due to post-translational modifications or degradation. However CyDye DIGE two-dimensional PAGE could be good methods to solve some of problems and more technologies should be explored to improve studies of proteins on multiple 2D SDS PAGE. For some proteins, such as AspS, ProS, TufA, abundance ratio between exponential growth and stationary phase were found differently from previous studies. This could due to different amount of protein bound to AIX column in these two growth phases. To test it, proteins in flow through fractions could be run on 2D PAGE

and specific protein could be identified by MS. However, even the amount were verified by 2D PAGE for proteins in both elution and flow-through, properties of same protein could be different for those bound to AIX column from those in flow-through. These properties could be enzyme activity, protein conformation, post-translational modification, protein-protein interaction, protein-molecule interaction, etc. Size exclusion column and native gel can be applied to study protein conformation. Mass spectrometry can be applied to study post-translation modification and protein-protein interaction. Different assays can test enzyme activity of specific protein.

Materials and methods

Sample preparation

Cultures of *E. coli* K-12 strain MG1655 were grown in 2-liter, non-baffled Fernbach flasks containing 1 liter of MOPS minimal medium (Neidhardt REF) (0.4% glucose, 19 mM NH_4Cl, 1.32 mM K_2HPO_4, 2 μg/ml thiamine, 10 μg/ml uridine, 0.52 mM $MgCl_2$, 0.25 μM $CaCl_2$, 8.37 g/l MOPS, 0.72 g/l tricine, 48 mg/l K_2SO_4, 2.92 g/l NaCl, 3 mg/l $FeSO_4 \cdot 7H_2O$, and micronutrients). Cultures were inoculated with 5 ml of an overnight culture grown in the same medium at 37°C with aeration for 16-18 hr. Cultures were grown with aeration (250 RPM) in a 37°C water bath. For the exponential phase sample, one culture was harvested when it reached OD600 ≅ 0.5. For the early stationary-phase sample, the other culture was harvested about 3 hr after the onset of stationary phase (OD600 ≅ 1.1). Lysates from about 5×10^8 cells were prepared as described in chapter II, materials and methods.

Anion exchange column-HPLC

Five ml of the clarified cell lysates, containing ~30 mg of protein as estimated by Bradford assay, were loaded onto a 1 ml AIX (Q15) 5/50 column (Amersham Biosciences) on an AKTA Purifier system (Amersham Biosciences). After washing the column, proteins were eluted a 50 mM to 1M NaCl gradient with a flow rate of 3ml/min (Champion *et al.*, 2003) and collected in 5 ml fractions. Protein was detected only fractions 5 through 23 by Bradford assay.

Two-dimensional SDS polyacrylamide gel electrophoresis

Two-dimensional PAGE was performed by the Protein Chemistry Laboratory at Texas A & M University. For the exponential phase sample, 1 ml of each protein-containing fraction was precipitated by TCA. TCA was added to a final concentration of 14%. To have the same amount of protein from the stationary phase fractions, 740 µl of each protein-containing fraction was precipitated by 14% TCA. Before precipitation, 10 µg of horse myoglobin was added to each fraction as an internal marker of protein abundance. Horse myoglobin was chosen as an internal marker because it is not an *E. coli* protein and will not affect the analysis of *E. coli* proteins, and it has a low molecular weight (17.6 kD) and high pI (7.3) so it will migrate at the right lower corner of the 2D gels and not overlay any *E. coli* proteins. TCA precipitated samples were washed twice with cold acetone and air dried. Samples were resuspended and focused on Igphor immobilized gradient gels (14 cm pH 3–10 Nonlinear, Pharmacia) for 60–80,000 Volt hours. After reduction and exchange in SDS and DTT, 12% polyacrylamide gels (14 x 14

cm) were run in the second dimension. The gels were stained with Coomassie blue R-250 to visualize protein spots. The 2D PAGE were scanned, Gel images are stored in the web-based database. Gel images for the exponential growth sample are located at http://eep.tamu.edu/nondelc/results/log750.php. Gel images for the stationary phase sample are located at http://eep.tamu.edu/nondelc/results/stat750.php.

Analysis of 2D PAGE

Gel images of 2D PAGE were analyzed using the Z3 2D gel analysis system. The area of each visualized spot is circled. The Z3 software assigns an ID number, locates the coordinates, and calculates the intensity of each spot. Proteins with post-translational modifications could have multiple spots at different locations on a gel; each of these spots would be identified as separate proteins and assigned a unique ID.

By comparing the intensity of each spot with the intensity of the internal marker myoglobin, the amount of each protein was calculated. Spots with amount lower than 100 ng, the lowest amount detected by coomassie blue, were filtered out. The Z3 software assigns coordinates using units of pixels. To get the real location of spots on the 2D gels, the coordinates were changed from pixels to centimeters. With the coordinates in cm, an observed pI and observed MW was calculated based on the molecular weight markers run in the second dimension and the known pI gradient (Pharmacia) for the first dimension

To compare gels from the two growth phases, the 2D gels from adjacent fractions of the same growth phase were compared and linked, and then compared to the 2D gels of the corresponding fractions from the other growth phase. For example, 2D gels of

exponential phase fractions 5 and 6 were overlaid first. Then the 2D gels of fractions 5 for both exponential and stationary phase samples were overlaid. Visualized spots were assigned ID numbers, coordinates and spot intensities. Spots at the same coordinates on all three gels were assigned the same ID. Next, the 2D gels of stationary phase fractions 5 and 6 were overlaid. As a result, spots on fractions 5 and 6 were linked. Then same procedure was done for fractions 6 and 7, 7 and 8, etc. In this way, the proteins detected on 2D gels of fractions 5 through 23 from both growth phases were linked together.

Montage in-gel digestion

Coomassie-stained spots were excised from the 2D PAGE. In-gel digestions were carried out using the Montage In-Gel Digest Kit (Millipore).

Tandem mass spectrometry

Tandem mass spectrometry was performed as described in chapter II, materials and methods.

Methods to generate virtual 2D gels and redraw real 2D gels

To generate virtual 2D gels for every fraction from AIX column for NonDeLC study, spot for every protein was plotted on two-dimensional images using the calculated pI and calculated MW of every protein. To redraw real 2D gels, every visualized spot on the real 2D gels were re-plotted on two-dimensional images using the observed pI and observed MW of every spots. The first dimension is for pI and the second dimension is for MW. This work was done using Perl script.

CHAPTER IV

EEP

Background

Studies of genomics, transcriptomics, metabolomics, proteomics and other 'OMICS' create large amount of data, which are very hard to organize and analyze either manually or by using spreadsheets such as Excel. Relational Database Management Systems (RDBMS) are more efficient tools to store, organize and analyze data. Moreover, scripts can be written to query such databases and to generate a web-based interface to allow access to data and analysis by multiple users over the Internet. In this chapter, I describe the construction of Experiments in *E. coli* Proteomics (EEP), a web-based RDBMS I wrote to facilitate analysis of proteomics experiments done in our lab, including the non-denaturing liquid chromatography (NonDeLC) experiments described in dissertation of Matthew Champion, the Heparome experiments described in chapter II, the 2D PAGE experiments from chapter III. In addition, EEP was designed to allow incorporation of data from proteomic studies by other groups for comparative analysis. EEP contains information from SwissProt 2D PAGE proteomics (Tonella *et al.*, 2001), the CyberCell 2D PAGE project (Sundararaj *et al.*, 2004) and Corbin's 2D-LC proteomics (Corbin *et al.*, 2003). The history of any protein in these proteomic studies can be searched through the EEP databases.

Results

Database and website structure

To construct databases for biological studies, the content of the database should contain not only the data of final results but also description of the experiment procedure for each experiment, because minor difference of experiments procedure could cause big difference to results. Based on the experimental design and procedure, a logical structure of database could be constructed by creating multiple tables that describe the details of every step of experiments. Relational databases allow representation of logical relationships between tables such as one to one, one to many or many to many, which are more difficult to represent in spreadsheets.

Databases in EEP were constructed using MySQL (http://www.mysql.com/) and Perl (http://www.perl.com/) to manage the information about proteins identified in different experiments. Different databases were constructed for the Heparome, NonDeLC and 2D PAGE studies. EEP unifies the querying of these three databases. In the database Heparome, tables for experiments, chromatography, fractions, gels, bands, lc_ms, ms and msms were created. Different tables store information from different steps of the experiments (Fig. 4.1A). A similar database was also constructed for NonDeLC (Fig. 4.1B). Information from proteomics studies from other groups is also store in the NonDeLC database. Right now only 2 tables were constructed for 2D PAGE. Information from tables in different databases is sharable, such as information in table ecogene and table nc_000913, which store annotations of whole *E. coli* genome for MG1655.

To display the databases to other people, a website interface was built using PHP (http://www.php.net/), Perl-CGI and HTML (Fig. 4.2).

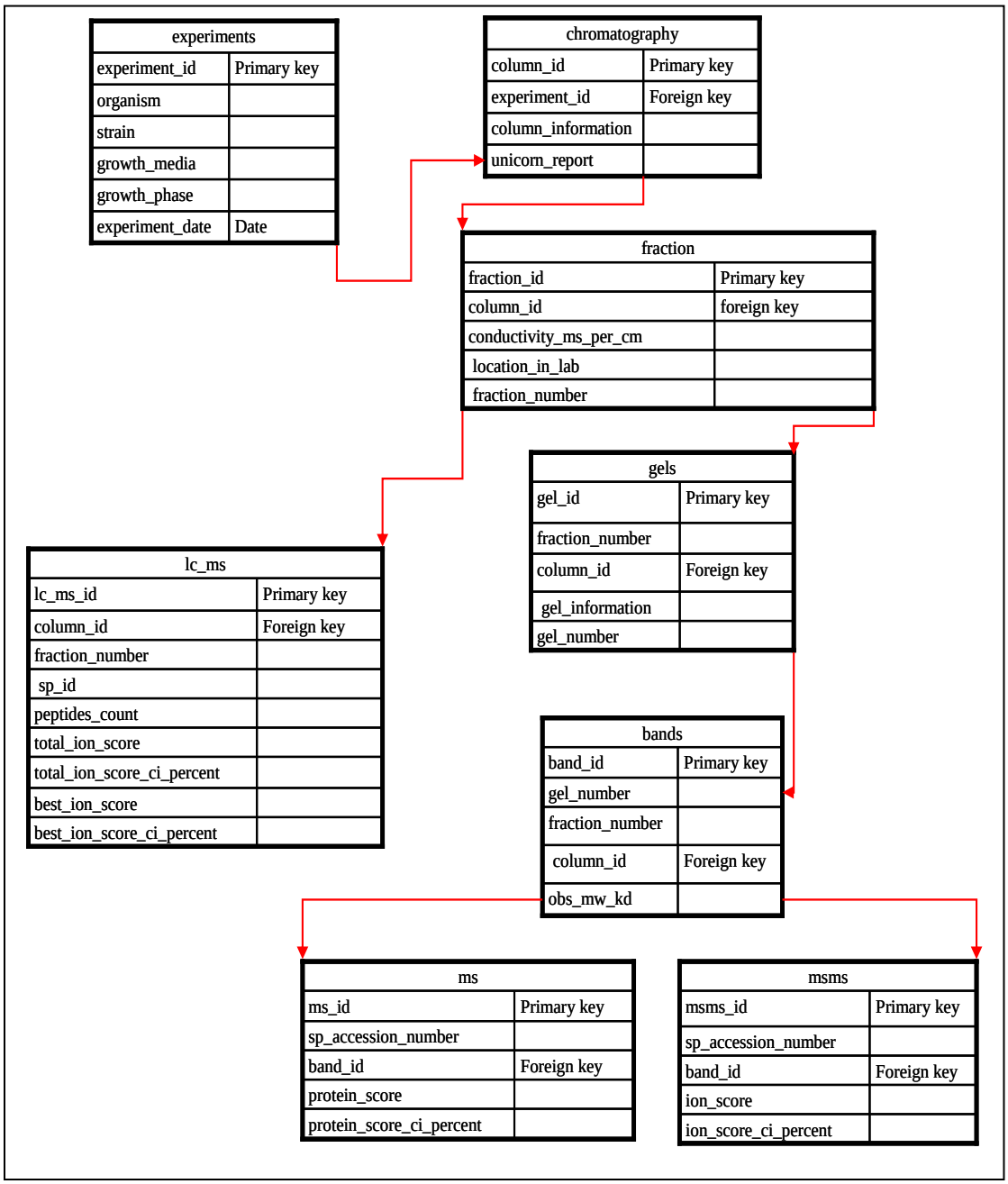


Figure 4.1A. Structure of the MySQL database of heparin-binding proteins (Heparome).

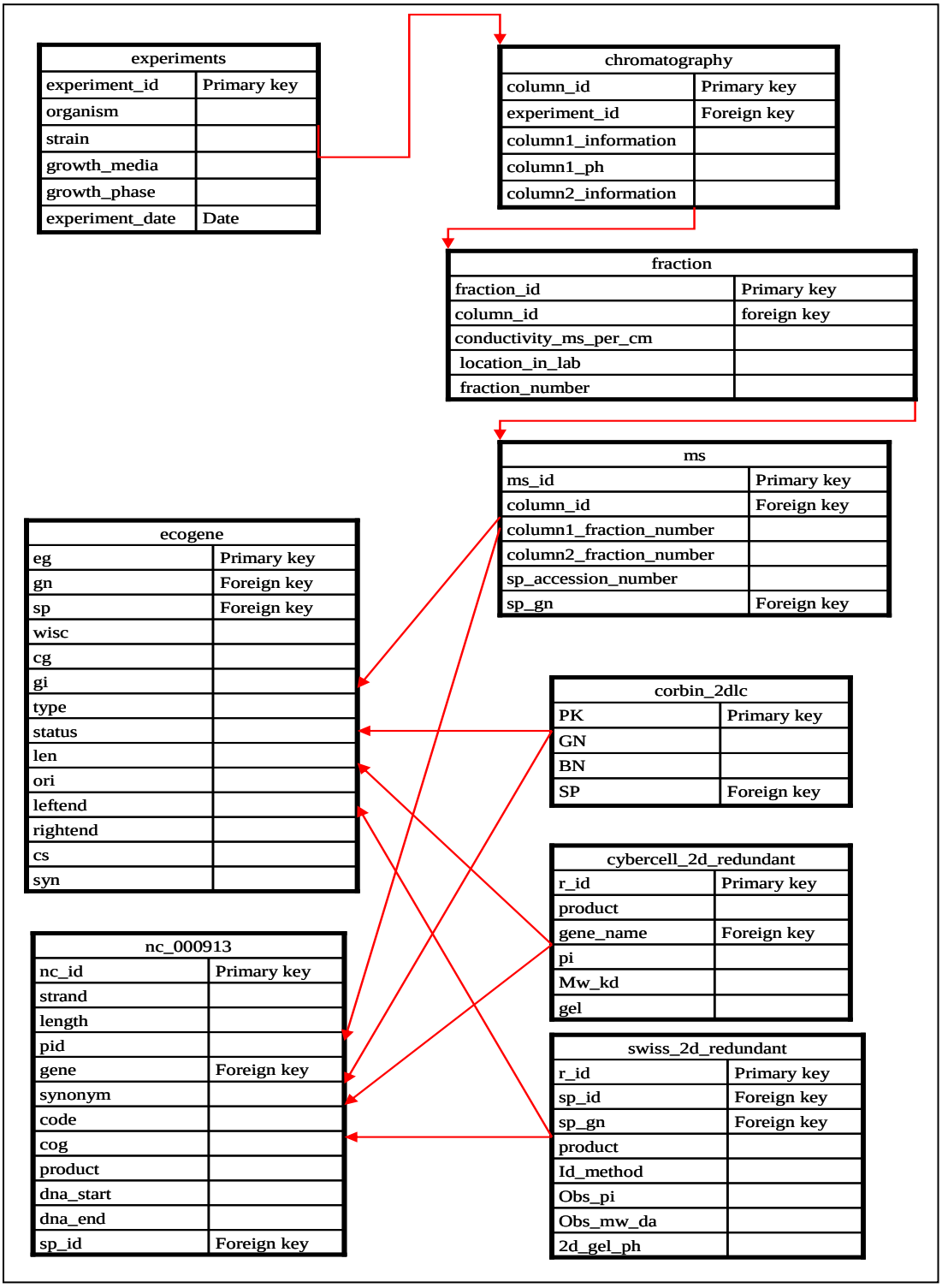


Figure 4.1B. Structure of the MySQL database of NonDeLC and other proteomic studies.

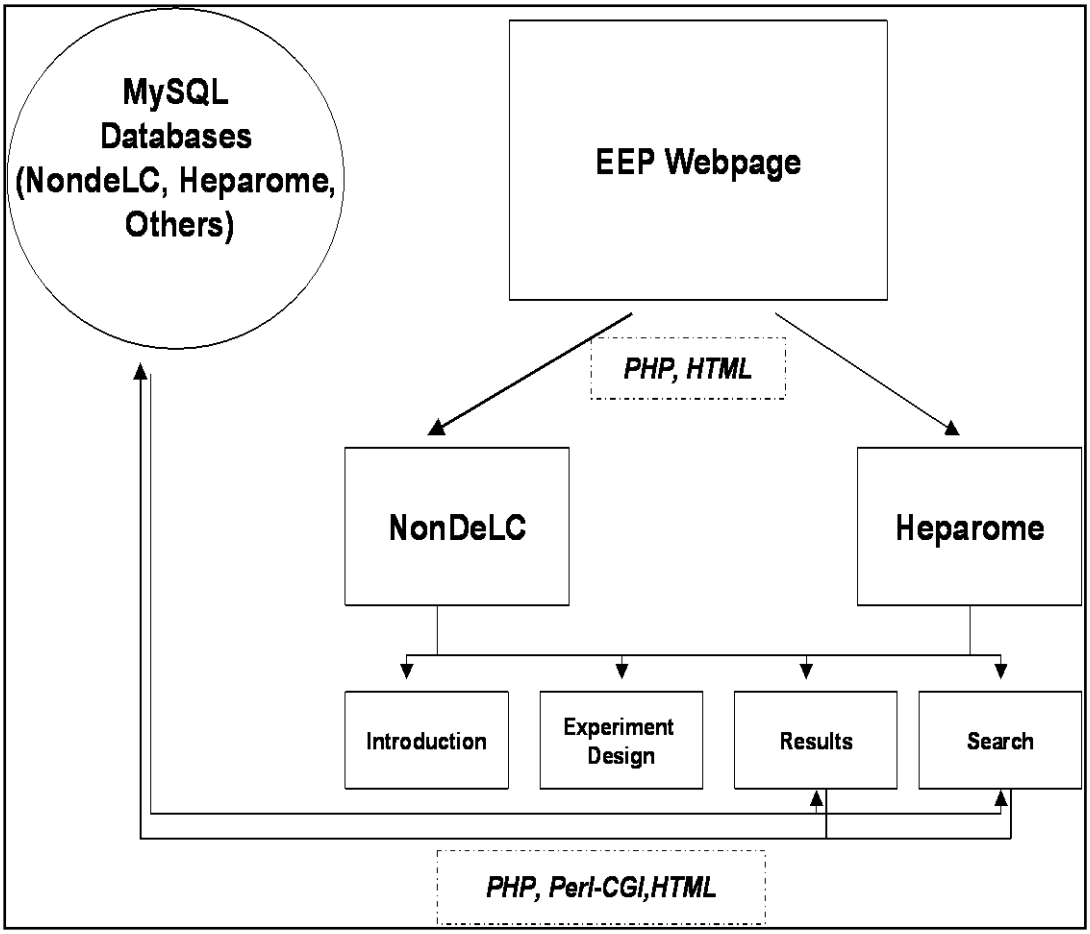


Figure 4.2. Connections between the webpage interface and the MySQL databases. Circles or boxes with solid lines are elements in database and website. Boxes with dashed lines are languages used to create the connection between different elements.

Website content

In the website for NonDeLC, Heparome or 2D PAGE, there are basically 4 sections: introduction, experiment design, results and search (Fig. 4.3). Content in the introduction and experiment design sections was generated as static HTML. Queries from results and search were dynamically generated using Perl-CGI and PHP. Section 'About' contains an introduction to the project, background, methods and people who worked on the project. Section 'Experiment design' describes the strategies of experiments. Section 'Results' contains details of experiment results, such as proteins identified in each growth phase or in every fraction, proteins up-regulated in certain growth phase, etc. Section 'Search' contains frequently used queries, such as elution pattern of every protein, or comparison of elution patterns between any two proteins on NonDeLC.

Fig 4.4 shows an example of the visualization, using DapA and DapB (Fig. 4.4). By looking at the visualized elution patterns, DapA had different elution positions between exponential growth and stationary phase, while DapB had similar elution positions between these two growth phases. Also, DapA and DapB did not elute out together in either growth phase. Many displays in EEP also retrieve basic information about identified proteins (Fig. 4.5), such as gene name, SwissProt ID, b number, synonym(s), isoelectric point (PI), molecular weight (MW) and protein name. Gene name, SwissProt ID, b number, synonym(s) and proteins name were obtained from EcoGene (Rudd, 2000). Isoelectric point (pI) and molecular weight (MW) were obtained from SwissProt (Tonella *et al.*, 2001). Also links to other proteomic databases are provided, such as NCBI, Expasy Proteomics Server, GenProEC, EcoCyc and Cyber Cell Project. Details

of experiments and database construction are posted on the website also.

Database access

EEP can be accessed at the website: http://eep.tamu.edu.

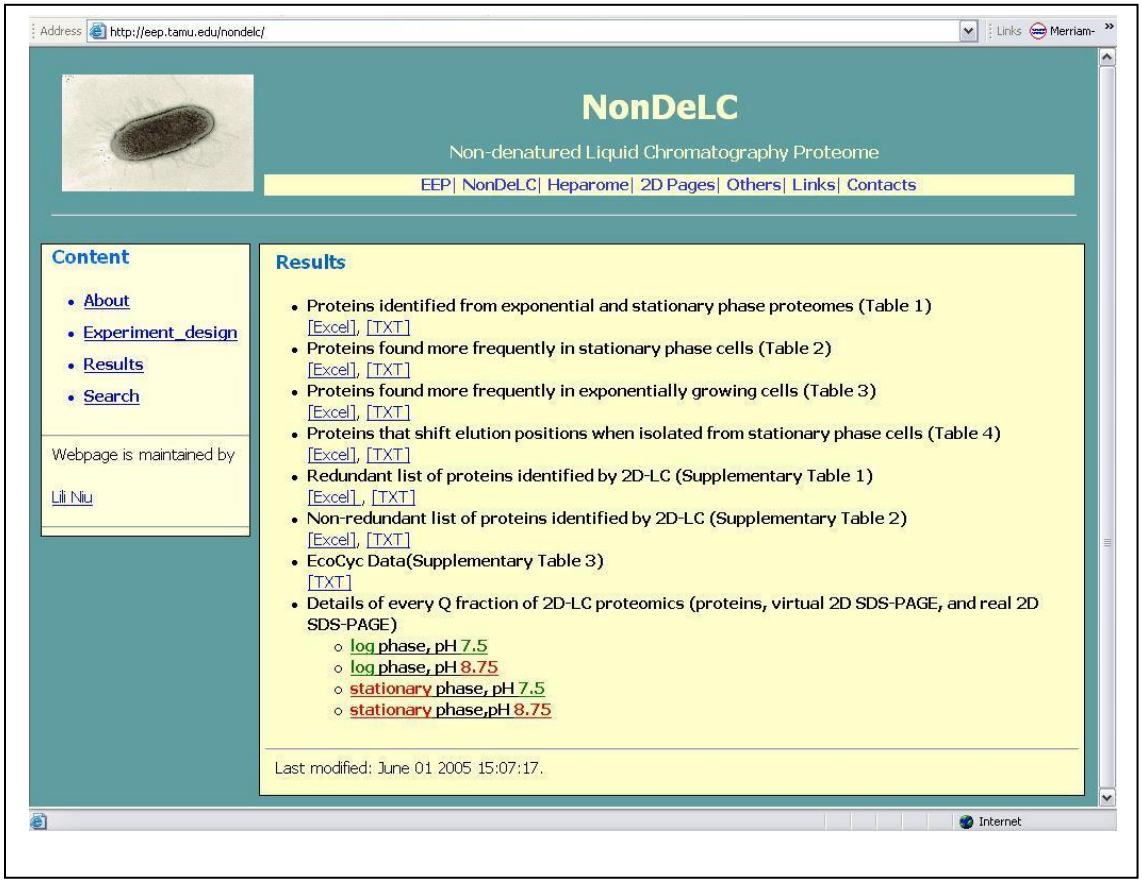


Figure 4.3. The website content for NonDeLC. The website contains 4 sections: About, Experiment design, Results, and Search.

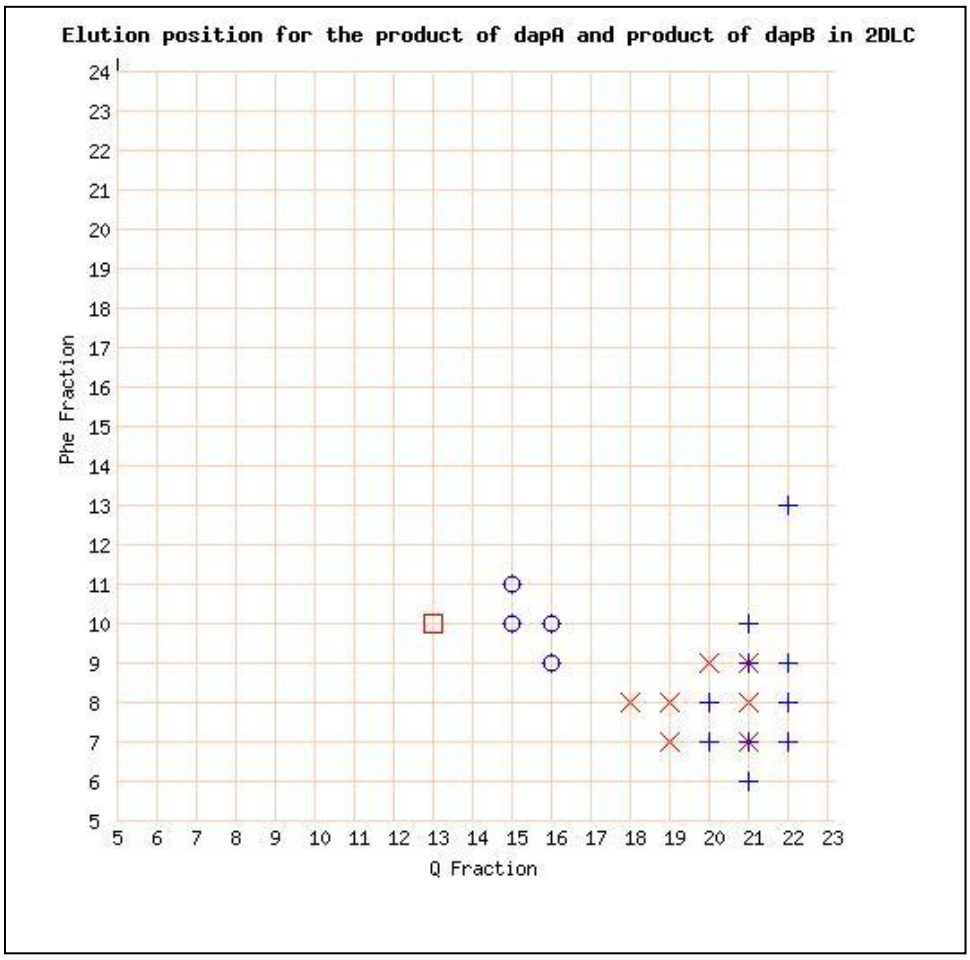


Figure 4.4. Comparison of elution patterns between DapA and DapB on NonDeLC. X-axis is the fraction number from anion exchange chromatography at pH 7.5, and Y-axis is the fraction number from hydrophobic interaction chromatography. O: Positions of DapA on NonDeLC from exponential growth sample. □: Positions of DapA on NonDeLC from stationary phase sample. +: Positions of DapB on NonDeLC from exponential growth sample. ×: Positions of DapB on NonDeLC from stationary phase sample.

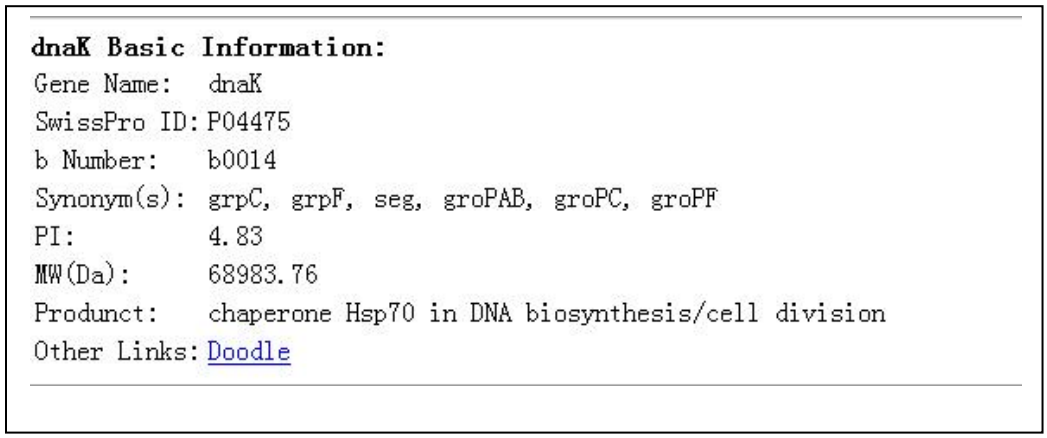


Figure 4.5. Information of DnaK identified by NonDeLC. pI is isoelectric point. MW is molecular weight.

Discussion

EEP provides an interface, which facilitates not only the data management for us, but also the data mining for the scientific community. Using tools of database and computational techniques, the studies from NonDeLC, Heparome and 2D SDS PAGE proteomic studies were combined and compared with proteomic studies by other groups.

Using GD, EEP provide a visualization tool to display the elution position of any identified protein. For example, in NonDeLC project, EEP provides a query (http://dimer.tamu.edu/cgi-bin/eep/elution_nondelc_Q_phe.pl) which can retrieve and visualize the elution positions of any identified protein from samples of both exponential growth and stationary phase. By looking at the visualized elution pattern of certain protein, one can easily tell if there is any difference of elution position between different growth phases. Proteins with difference of elution position could be candidates for study of physiology changes in stationary phase.

EEP can also help us look for protein complex. For example, in NonDeLC project, elution positions of any pair of proteins could be visually displayed together to facilitate the study of possible protein complex. By looking at the visualized elution positions of any two proteins, proteins could be candidates of protein complex if these two proteins eluted at same position(s), such as SucC and SucD in exponential phase (http://dimer.tamu.edu/cgi-bin/eep/elution_nondelc_Q_Phe_pair.pl?pH=7.5&item=gene_name&data=SucC&data2=SucD&submit_query=Search).

CHAPTER V

BIOINFORMATICS

Background

This chapter describes bioinformatics tools that were constructed during the course of this dissertation research. These were focused in two areas. First, I improved the web-based database, Doodle (http://dimer.tamu.edu/doodle), which was constructed by Leonardo Mariño-Ramírez. This included updating *E. coli* genome annotation and modifying how we display different aspects of the genome annotation of *E. coli*, using the GMOD genome browser, Gbrowse. Second, I developed bioinformatics tools for the Practical Genomics for Undergraduates Project (PGU) (http://dimer.tamu.edu/young/genomics/) run by Ryland Young and Elizabeth Summer. PGU sequences and annotates *Burkholderia* phage genomes. With more and more phage genomes completed, manual annotation of the genome sequence becomes a significant burden. Although no genome can be accurately and comprehensively annotated by automated methods alone, computational programs can be useful tools for making manual genome annotation easier. Phage and bacteria present special challenges for bioinformatics analysis of genomes. To facilitate the PGU, I wrote several programs that are now used by the project. Some of these, such as “GetORF”, “LipoRy”, “Primary structure with charges”, and “Amino acid content”, were written because similar existing programs were not easily obtainable or needed to be modified for specific requirements, including web access. Other programs, such as “GetRzRz1”, “SAREndolysin”, and

"Phage lysis genes", were written to search and predict specific phage genes. The program "GroupProtein" was written to group phage genes into families.

Results

Programs for E. coli genome annotation

Workflow for updating Doodle

Doodle is an acronym for the Database of Oligomerization Domains from Lambda Experiments. Doodle was built primarily to serve as a laboratory information management system to track the clones, sequences, and analysis of self-assembling domains from repressor fusion cloning projects (Marino-Ramirez and Hu, 2002; Marino-Ramirez *et al.*, 2003). However, Doodle also serves as a source of information about *E. coli* proteins based on information compiled from other resources and from analysis done locally. The structure and pipeline of Doodle database were described in the dissertation of Leonardo Mariño-Ramírez and also in the web-accessible documentation (http://dimer.tamu.edu/doodle/docs/intro.html). In this section I describe how I improved the updating of information in Doodle.

Updating the data in Doodle involves three kinds of operations: 1) adding new interacting sequence tag (IST) annotations (this did not occur in my work), 2) updating information from other sources, such as identifying proteins with Protein Data Bank (PDB) homologs, proteins with InterPro hits, and 3) correcting the coordinates of features in the database, based on changes in the reference sequence.

To update information in the Doodle database, several tables needed to change to

stay synchronized with their external data sources. Basically, most of the tables, such as "gi2gene_name", "ist", "pdb_feature", "coils", and "spinterpro", need to be updated every time the genome sequence or annotation changes for *E. coli* K-12 MG1655. Due to changes in the genome sequence and annotation, basic information, such as coordinates and sequence for DNA and protein, gi number, b number and products, in relevant tables need to be updated.

Because new X-ray structures are being deposited in the PDB at a rapid rate, we periodically rescan the PDB for *E. coli* proteins and homologs. To do this, the pdbaa file is obtained from PBD (http://www.pdb.org/). To update the table "pdb_feature" in the database, which is used to generate lists of PDB homologs (http://dimer.tamu.edu/cgi-bin/doodle/homo.pl?org=E), the script "genome2blast" performs a BLAST of all *E. coli* proteins against all proteins in the PDB, the script "parse_pdb" parses the BLAST result, and the script "pdb2feature" uploads parsed BLAST results into the database.

To update proteins with InterPro hits, which are protein families, domains and functional sites, the protein2interpro.dat.gz file is obtained from InterPro database (ftp://ftp.ebi.ac.uk/pub/databases/interpro/). To update the table "spinterpro", which is for proteins with InterPro hits, the script "swiss2ip" is run to generate an interpro file for *E. coli*, and the script "interpro2doodle" is run to upload data into database.

To correct the coordinates of features in the database based on changes in the reference sequence, the tables "gi2gene_name", "ist", "pdb_feature", "coils", and "spinterpro" need to be updated. Processing and updating tables "pdb_feature" and "spinterpro" have been mentioned above. Table "gi2gene_name" stores the genome

annotation of the *E. coli* strain MG1655. To update information in table "gi2gene_name", the NC_000913.ptt file, which contains the annotation of *E. coli* K-12 MG1655, was obtained from NCBI (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia_coli_K12/NC_000913.ptt). The ECOLI.dat file was obtained from Expasy (ftp://ca.expasy.org/databases/complete_proteomes/entries/bacteria/ECOLI.dat). The script "NC_SP2gi2genename" regenerates the table "gi2gene_name" using the files NC_000913.ppt and ECOLI.dat. The script "gi2gene_name_loader" uploads the new "gi2gene_name" table into the database.

To update coordinates for features in the table "ist", sequences of ISTs are compared to the new whole genome DNA sequence of *E. coli* MG1655 using BLAST by the script "dir2blast". The script "parse_report" parses the blast result to a form that is uploaded into the database using the script "txt2feature".

To update coordinates feature in table "coils", which store the coiled coil predictions for the whole genome, the script "fasta2ccp" is run to re-predict all the coiled coils encoded in the annotated proteins in the genome using the COILS2 program of Lupas (Lupas, 1997). Results are parsed by the script "coils2pos.pl" and uploaded into database by the script "coils2doodle".

Gbrowse for E. coli genome annotation

Graphical genome browsers are used to display annotations in their genome context. Gbrowse (Stein *et al.*, 2002) is a powerful, open source genome browser and is part of the Generic Model Organism Database project (http://www.gmod.org). Gbrowse is a

web-based tool developed by the group of L. Stein. It has been utilized by different studies such as the Saccharomyces Genome Database (Nash *et al.*, 2007), Wormbase (Harris *et al.*, 2003), SwissRegulon (Pachkov *et al.*, 2007), PseudoCAP (Winsor *et al.*, 2005) and soybean genome database (Shultz *et al.*, 2006), and many others.

To facilitate the display of genome annotations for *E. coli* and its plasmids and phages, we installed and customized a local installation of Gbrowse version 1.64 at http://dimer.tamu.edu/cgi-bin/gb/gbrowse. Fig.5.1 shows a screenshot for display of *E. coli* annotations for strain K-12 MG1655. The display of annotations in Gbrowse is based on annotation tracks, which can be toggled to display or hide a particular set of annotations. Table 5.1 shows annotation tracks for our *E. coli* Gbrowse. There are also separate Gbrowse database for strain K-12 W3110 (Fig. 5.2) and for 28 bacteriophages. To generate different tracks into Gbrowse, a Gene Finding Format (GFF) (http://www.sanger.ac.uk/Software/formats/GFF/) file for every track was created and loaded into the Gbrowse database. GFF is a format for describing genes and other features associated with DNA, RNA and Protein sequences. In Gbrowse version 1.64, GFF2 was used. Also the information of different tracks need to be added into the configure file for each specific strain or phage.

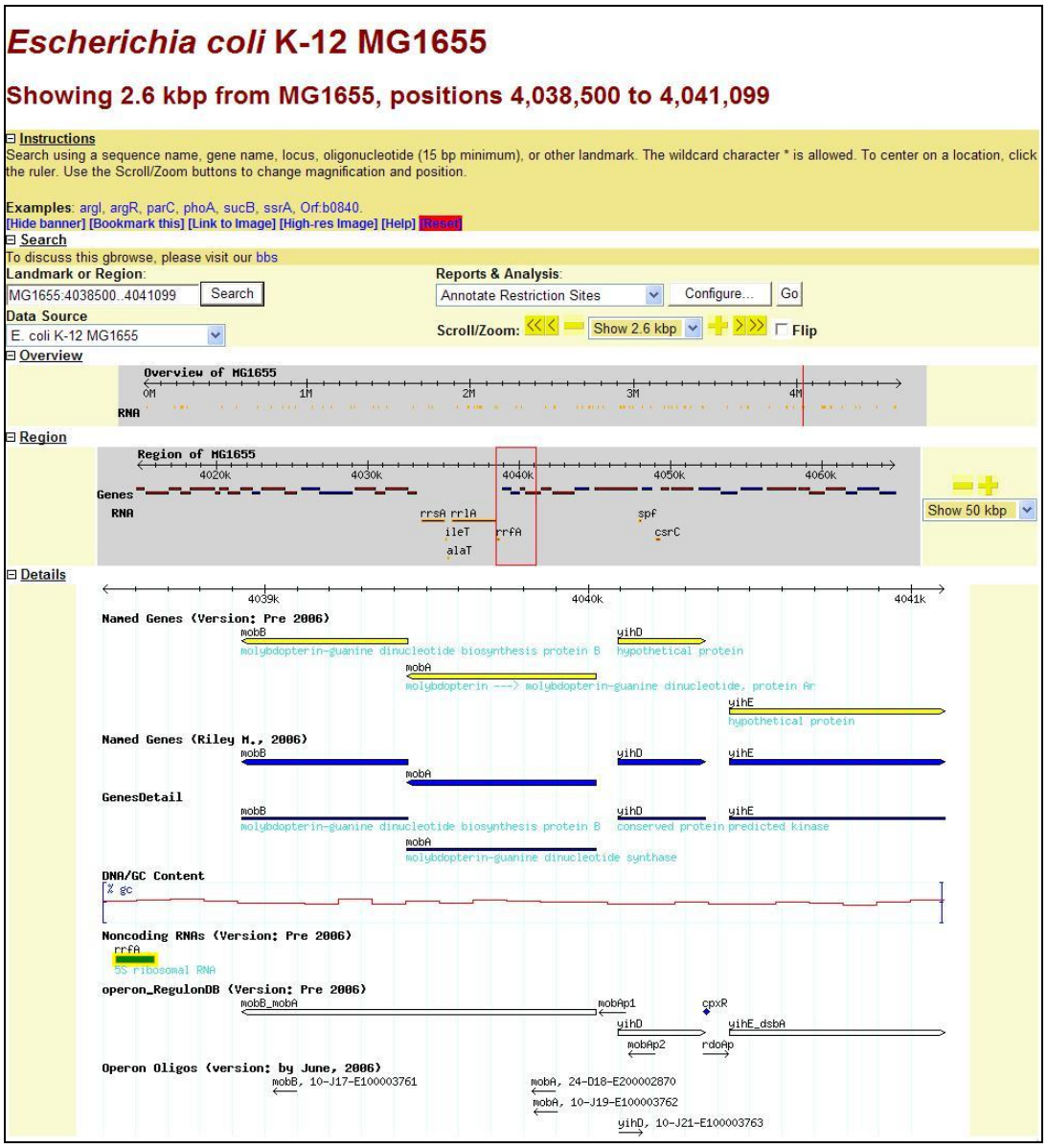


Figure 5.1. The genome view of *E. coli* K-12 MG1655 in Gbrowse. Information for two versions of annotation (U00096.1 and U00096.2), DNA/GC content, noncoding RNAs, operons from RegulonDB (operon, promoter, TF binding site), and information for oligos from company Operon were included.

Table 5.1. Annotation tracks for *E. coli* Gbrowse

Annotation tracks	Strain (s) applied
Named Genes (Version: Pre 2006)	K-12 MG1655 and K-12 W3110
Named Genes (Riley *et al.*, 2006)	K-12 MG1655 and K-12 W3110
Genes	K-12 MG1655
GenesDetail	K-12 MG1655
ORF (Version: Pre 2006)	K-12 MG1655
IST	K-12 MG1655
SwissProt (Version: Pre 2006)	K-12 MG1655
COG (Version: Pre 2006)	K-12 MG1655
3-frame translation (forward)	K-12 MG1655 and K-12 W3110
3-frame translation (reverse)	K-12 MG1655 and K-12 W3110
DNA/GC Content	K-12 MG1655 and K-12 W3110
Noncoding RNAs (Version: Pre 2006)	K-12 MG1655
Operon_RegulonDB (Version: Pre 2006)	K-12 MG1655
Operon Oligos (version: by June, 2006)	K-12 MG1655

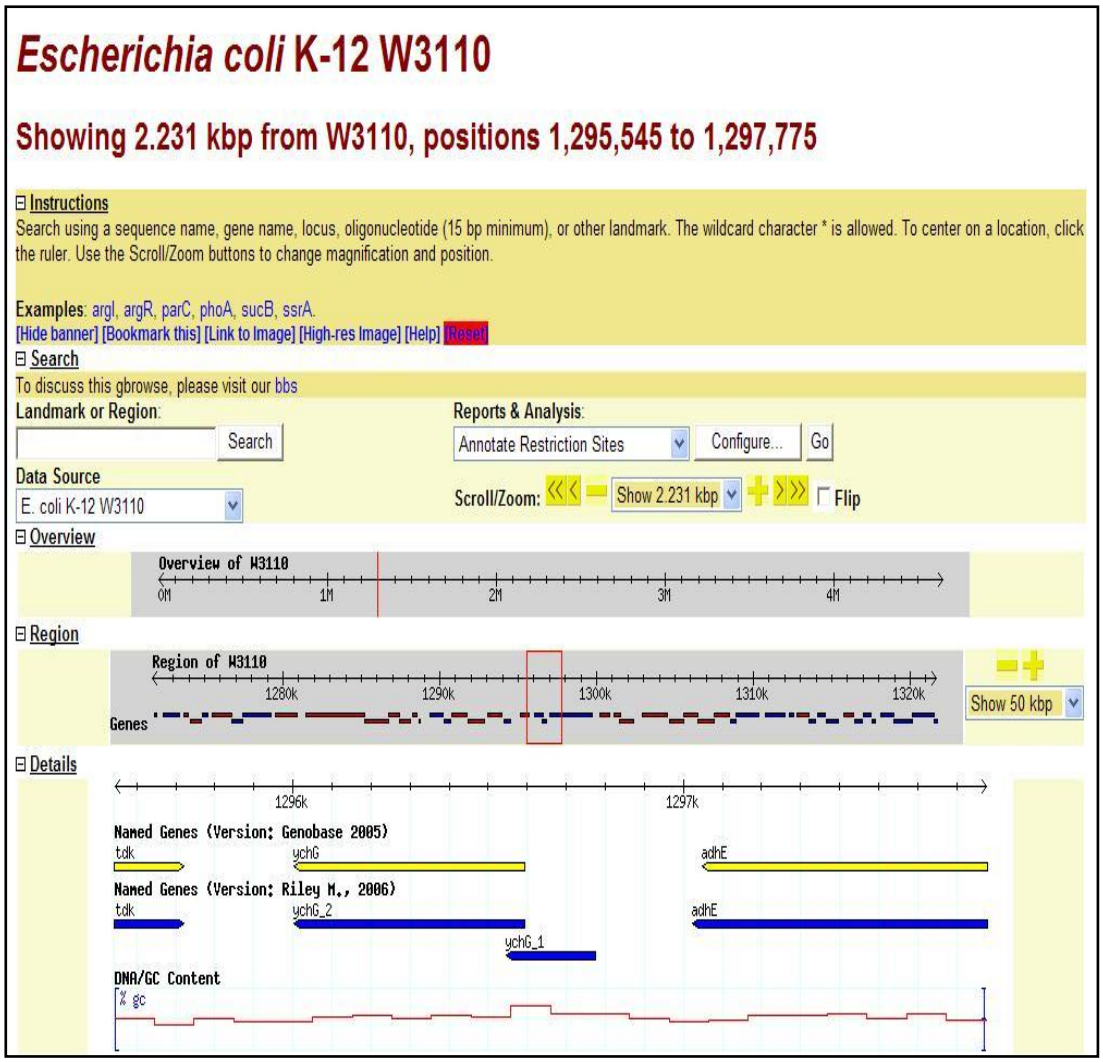
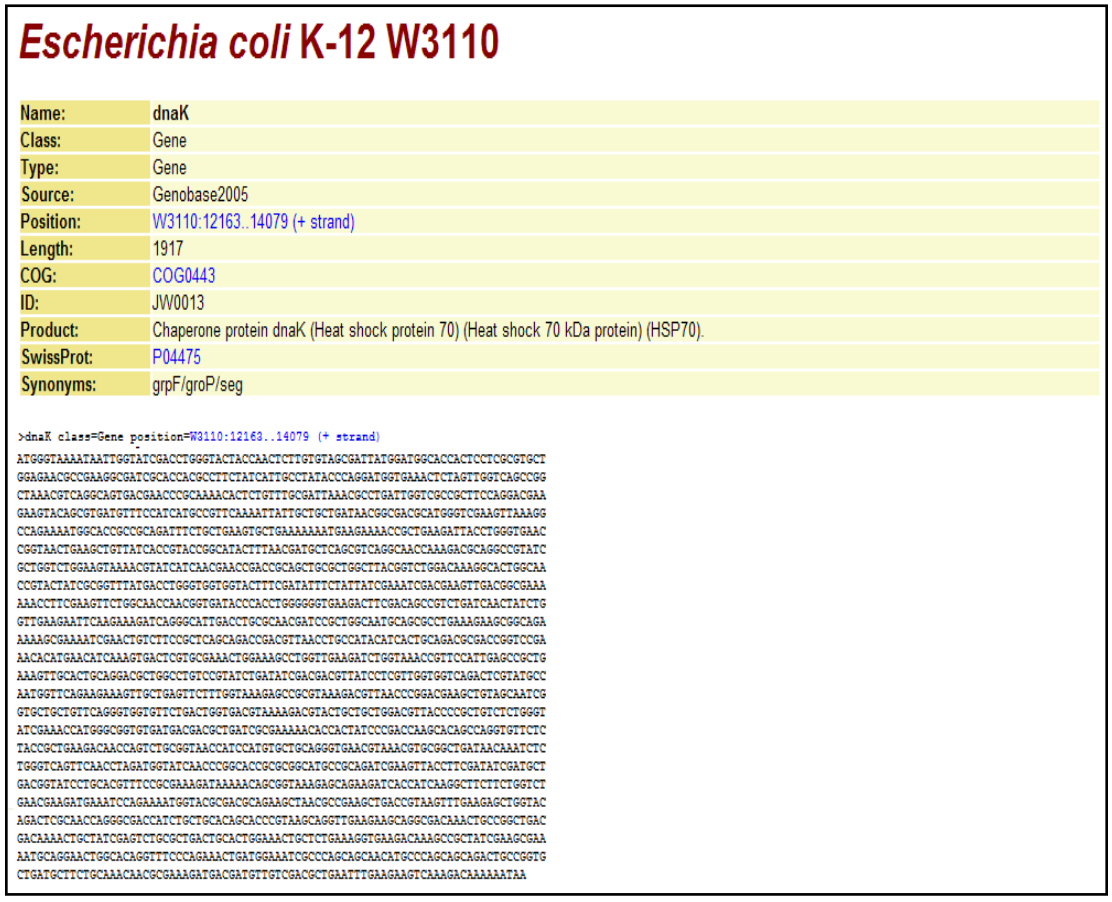


Figure 5.2. The genome view of *E. coli* K-12 W3110 in Gbrowse. Information for two versions of annotation (Genobase 2006 and Riley, Abe *et. al.* 2006) and DNA/GC content were included. In version of Riley M. 2006, *ychG_1* is a new gene and *adhE* has different coordinates from the previous annotation.

To generate the GFF file for every track, the existing annotation, I wrote Perl scripts to parse several *E. coli* annotation types into GFF format suitable for Gbrowse. Annotations of *E. coli* genome and RNA coding genes were obtained from NCBI (http://www.ncbi.nlm.nih.gov/). Annotation of operons was obtained from RegulonDB (http://regulondb.ccg.unam.mx/index.html). The position of oligonucleotide probes was obtained from Operon Company (http://www.operon.com). Since the genome sequences of both *E. coli* K12 MG1655 and W3110 were updated (Hayashi *et al.*, 2006), coordinates of genes from the previous annotation needed to be updated. First, gene sequences from previous annotation were retrieved from the old genome sequence using the script "fastacmd", which is part of the NCBI C toolkit (http://www.ncbi.nlm.nih.gov/IEB/ToolBox/SDKDOCS/INDEX.HTML). New coordinates for every gene were obtained by BLASTing the old gene sequences against the new genome sequence. As a result, gene coordinates obtained to populate a GFF file are corresponding to the gene models for the new genome sequence. For some annotations, a Perl script (ldas_bulk_load.pl) from the Gbrowse distribution was also used to load GFF files into an underlying MySQL database.

By clicking links of figures, details can be displayed in a Gbrowse details page. The Gbrowse details page contains information about the gene name, source, coordinates, gene length, COG number and link, JW ID, products name, SwissProt ID, synonyms and DNA sequence (Fig. 5.3). I configured our Gbrowse to allow searching genes not only with standard names, but also with ambiguous names, which information was obtained from EcoCyc (http://ecocyc.org/). For example both *dnaJ* and *dnaB* have *groP* as a

synonym. By searching *groP*, these two genes are listed and a summary of each one is displayed, including all synonyms and gene coordinates (Fig. 5.4).



Escherichia coli K-12 W3110

Name:	dnaK
Class:	Gene
Type:	Gene
Source:	Genobase2005
Position:	W3110:12163..14079 (+ strand)
Length:	1917
COG:	COG0443
ID:	JW0013
Product:	Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
SwissProt:	P04475
Synonyms:	grpF/groP/seg

>dnaK class=Gene position=W3110:12163..14079 (+ strand)

ATGGGTAAAATAATTGGTATCGACCTGGGTACTACCAACTCTTGTGTAGCGATTATGGATGGCACCACTCCTCGCGTGCT
GGAGAACGCCGAAGGCGATCGCACCACGCCTTCTATCATTGCCTATACCCAGGATGGTGAAACTCTAGTTGGTCAGCCGG
CTAAACGTCAGGCAGTGACGAACCCGCAAAACACTCTGTTTGCGATTAAACGCCTGATTGGTCGCCGCTTCCAGGACGAA
GAAGTACAGCGTGATGTTTCCATCATGCCGTTCAAAATTATTGCTGCTGATAACGGCGACGCATGGGTCGAAGTTAAAGG
CCAGAAAATGGCACCGCCGCAGATTTCTGCTGAAGTGCTGAAAAAAATGAAGAAAACCGCTGAAGATTACCTGGGTGAAC
CGGTAACTGAAGCTGTTATCACCGTACCGGCATACTTTAACGATGCTCAGCGTCAGGCAACCAAAGACGCAGGCCGTATC
GCTGGTCTGGAAGTAAAACGTATCATCAACGAACCGACCGCAGCTGCGCTGGCTTACGGTCTGGACAAAGGCACTGGCAA
CCGTACTATCGCGGTTTATGACCTGGGTGGTGGTACTTTCGATATTTCTATTATCGAAATCGACGAAGTTGACGGCGAAA
AAACCTTCGAAGTTCTGGCAACCAACGGTGATACCCACCTGGGGGGTGAAGACTTCGACAGCCGTCTGATCAACTATCTG
GTTGAAGAATTCAAGAAAGATCAGGGCATTGACCTGCGCAACGATCCGCTGGCAATGCAGCGCCTGAAAGAAGCGGCAGA
AAAAGCGAAAATCGAACTGTCTTCCGCTCAGCAGACCGACGTTAACCTGCCATACATCACTGCAGACGCGACCGGTCCGA
AACACATGAACATCAAAGTGACTCGTGCGAAACTGGAAAGCCTGGTTGAAGATCTGGTAAACCGTTCCATTGAGCCGCTG
AAAGTTGCACTGCAGGACGCTGGCCTGTCCGTATCTGATATCGACGACGTTATCCTCGTTGGTGGTCAGACTCGTATGCC
AATGGTTCAGAAGAAAGTTGCTGAGTTCTTTGGTAAAGAGCCGCGTAAAGACGTTAACCCGGACGAAGCTGTAGCAATCG
GTGCTGCTGTTCAGGGTGGTGTTCTGACTGGTGACGTAAAAGACGTACTGCTGCTGGACGTTACCCCGCTGTCTCTGGGT
ATCGAAACCATGGGCGGTGTGATGACGACGCTGATCGCGAAAAACACCACTATCCCGACCAAGCACAGCCAGGTGTTCTC
TACCGCTGAAGACAACCAGTCTGCGGTAACCATCCATGTGCTGCAGGGTGAACGTAAACGTGCGGCTGATAACAAATCTC
TGGGTCAGTTCAACCTAGATGGTATCAACCCGGCACCGCGCGGCATGCCGCAGATCGAAGTTACCTTCGATATCGATGCT
GACGGTATCCTGCACGTTTCCGCGAAAGATAAAAACAGCGGTAAAGAGCAGAAGATCACCATCAAGGCTTCTTCTGGTCT
GAACGAAGATGAAATCCAGAAAATGGTACGCGACGCAGAAGCTAACGCCGAAGCTGACCGTAAGTTTGAAGAGCTGGTAC
AGACTCGCAACCAGGGCGACCATCTGCTGCACAGCACCCGTAAGCAGGTTGAAGAAGCAGGCGACAAACTGCCGGCTGAC
GACAAAACTGCTATCGAGTCTGCGCTGACTGCACTGGAAACTGCTCTGAAAGGTGAAGACAAAGCCGCTATCGAAGCGAA
AATGCAGGAACTGGCACAGGTTTCCCAGAAACTGATGGAAATCGCCCAGCAGCAACATGCCCAGCAGCAGACTGCCGGTG
CTGATGCTTCTGCAAACAACGCGAAAGATGACGATGTTGTCGACGCTGAATTTGAAGAAGTCAAAGACAAAAAATAA

Figure 5.3. The Gbrowse details page for *dnaK* in version of Genobase 2005 of *E. coli* K-12 W3110. This page contains information of gene name, source, coordinates, gene length, COG number and link, JW ID, products name, SwissProt ID, synonyms and DNA sequence.

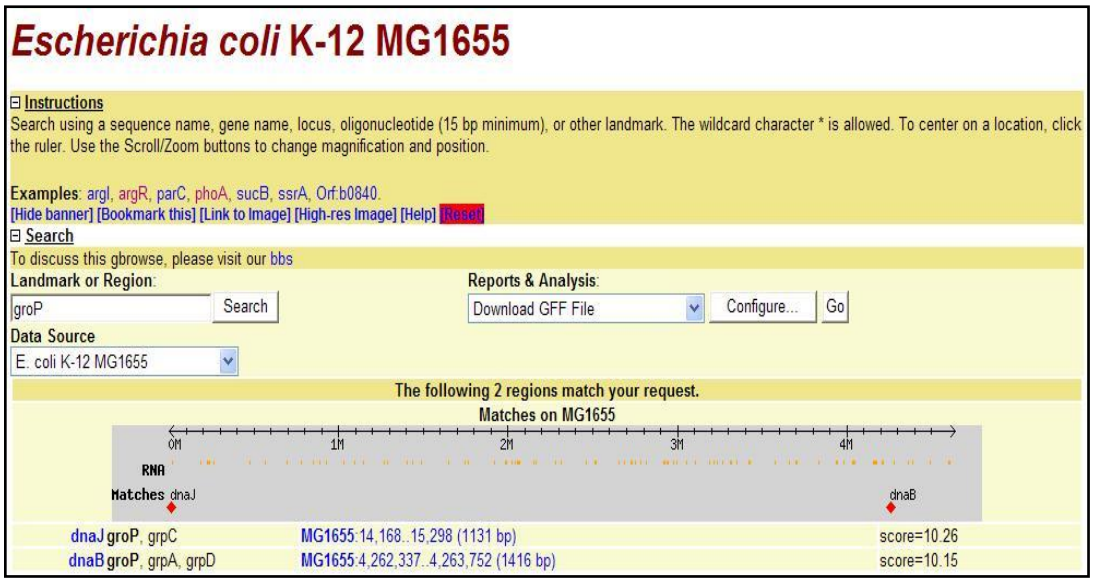


Figure 5.4. Genes with *groP* as synonyms in *E. coli* K-12 MG1655. Both *dnaJ* and *dnaB* have *groP* as a synonym.

Programs for the PGU project

With more and more *Burkholderia* phage genomes sequenced by the PGU project, annotation has become a major issue. To improve the process of annotation, computational programs were developed and applied.

GetORF

Many of the programs I wrote for the PGU involve identifying which polypeptide sequences encoded in a genome could be a member of a particular protein family. For example, "GetRzRz1" predicts genes encoding Rz and Rz1 genes. These programs apply rule-based filtering to a list of candidate proteins. Open reading frames (ORF) are sequences of DNA, which are located from a start codon to a stop codon that could be translated into proteins or polypeptides with no internal stop codons. The ORFs that correspond to actual genes are a subset of possible open reading frames in a genome. Part of the annotations of prokaryotic proteins are based on predictions of which ORFs in the DNA sequences correspond to the actual genes.

Many programs have been written for this purpose. One commonly used program is GeneMark.hmm (Lukashin and Borodovsky, 1998) (http://opal.biology.gatech.edu/GeneMark/heuristic_hmm2.cgi), which predicts genes from ORFs using hidden Markov models based on the properties of well-characterized genes. Although GeneMark gives the most confident gene candidates, which are often correlated to validated genes, for my purposes this filtering might lose too many potential gene candidates. For some predictions of specific genes, such as lipoproteins, the start positions of proteins are very important for prediction. It is better to have the

unfiltered information about all the possible ORFs. To predict all possible ORFs and translate them into proteins, GetORF (http://dimer.tamu.edu/cgi-bin/phage_genomics/get_orf.pl) was written. GetORF is used as a first step in many of the other programs described below.

By uploading the DNA sequence, choosing the start codon (ATG, TTG, CTG, or GTG), minimal gene length and strand, GetORF predicts all possible ORFs (Fig. 5.5). In the output, protein length, coordinates of genes and strand are displayed, under which protein sequences were also generated for every possible ORF, based on the codon usage of bacteria (Fig. 5.6).

Note that GetORF is not a program to predict genes. In other words, GetORF processes raw predictions that search all possible ORFs to facilitate further prediction of specific genes. Some programs in this chapter, such as LipoRy, utilize GetORF to get amino acid sequences of all possible ORFs and search proteins with unique features.

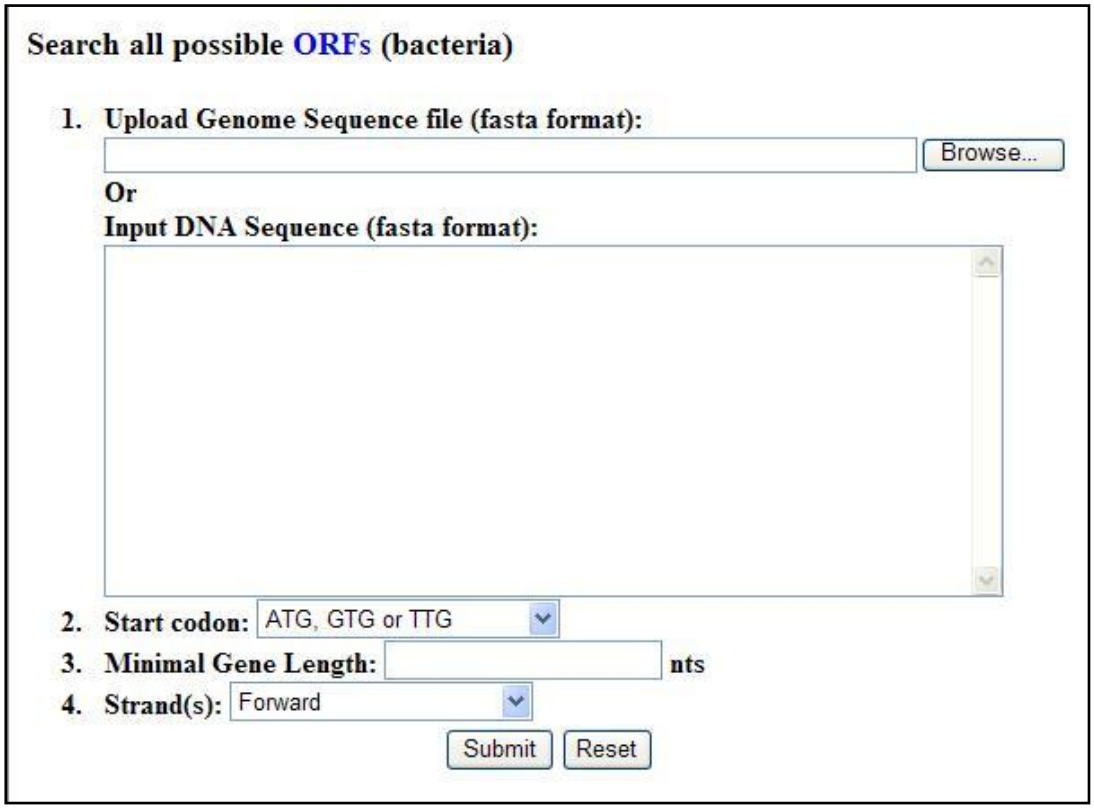


Figure 5.5. The form to submit ORF searching in GetORF.

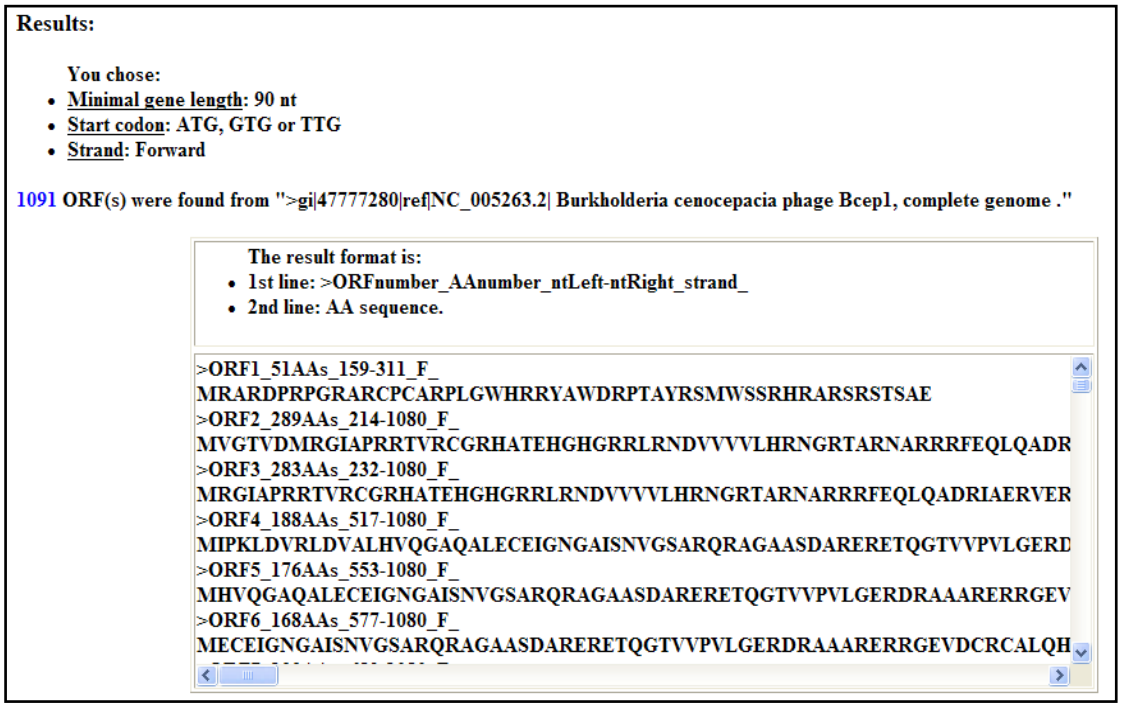


Figure 5.6. Results of ORF searching for the complete genome of *Burkholderia cenocepacia* phage Bcep1 by GetORF. 1091 ORF(s) were found for the forward (F) strands. Protein sequence, protein length, coordinates of genes and strand information were displayed.

LipoRy

Lipoproteins, containing both protein and lipid, play different roles, such as enzymes, transporters, structural proteins, antigens, adhesions and toxins. LipoP (Juncker *et al.*, 2003) has been widely used to predict lipoproteins of Gram-negative bacteria and their phages. However, LipoP is not well suited for the PGU project. LipoP is only available as a website; the source code is not downloadable. LipoP can process single protein sequences or multiple proteins as FASTA files, but it cannot predict lipoproteins using a DNA sequence. For new sequenced genomes, such as the phage genomes from the PGU project, without finished annotation, DNA sequence is a more accurate source for prediction. Also LipoP has difficulty performing bulk searches, larger than four thousand protein sequences over the web. Moreover, sometimes different formats for the output of the prediction are required for further annotation. Driven by the limitations of LipoP, LipoRy was written. LipoRy is based on rules generated by R. Young to predict lipoproteins from Gram-negative bacteria and their phages. All bacterial lipoproteins have a signal sequence in the N-terminal end followed by a cysteine (Hayashi and Wu, 1990). The signal sequence of lipoproteins is cleavable by lipoprotein signal peptidase (Lsp), which is also called signal peptidase II (SPaseII). The cysteine following a signal sequence is in the first 40 amino acids from N-terminal end, and the signal sequence contains at least 10 consecutive residues without charged amino acid. LipoRy is a web-based program (http://dimer.tamu.edu/cgi-bin/phage_genomics/Lipo_Ry.pl). Since the characters of lipoproteins are located at N-terminal end of proteins, the start positions of proteins are very important. Instead of

missing any possible candidates, GetORF is applied to predict all possible ORFs, from which candidates for lipoprotein are predicted. By uploading the DNA sequence or DNA FASTA file, choosing the start codon (ATG, GTG, CTG, and TTG), minimal gene length and strand (Fig. 5.7), candidates for lipoproteins will be predicted based on all possible ORFs (Fig. 5.8). Besides the basic information of ORFs, cysteines following the signal sequence are highlighted in yellow. Since the start source is DNA sequence and the ORF prediction search gives all possible ORFs without filtration, there will be higher percentage of false positives among the candidates from LipoRy than from LipoP. For example, LipoRy predicts 187 lipoprotein candidates for bacteriophage lambda. Out of these 187 candidates, 26 of them are also predicted by LipoP. However for some studies, such as the beginning of genome annotation for phages in PGU project, it is better to have false positives than false negatives. More restrictions could be added for more specific properties of certain genes to filter out more false positives. For example LipoRy was integrated into the program "GetRzRz1" to predict *Rz* and *Rz1* gene pairs in phage genomes.

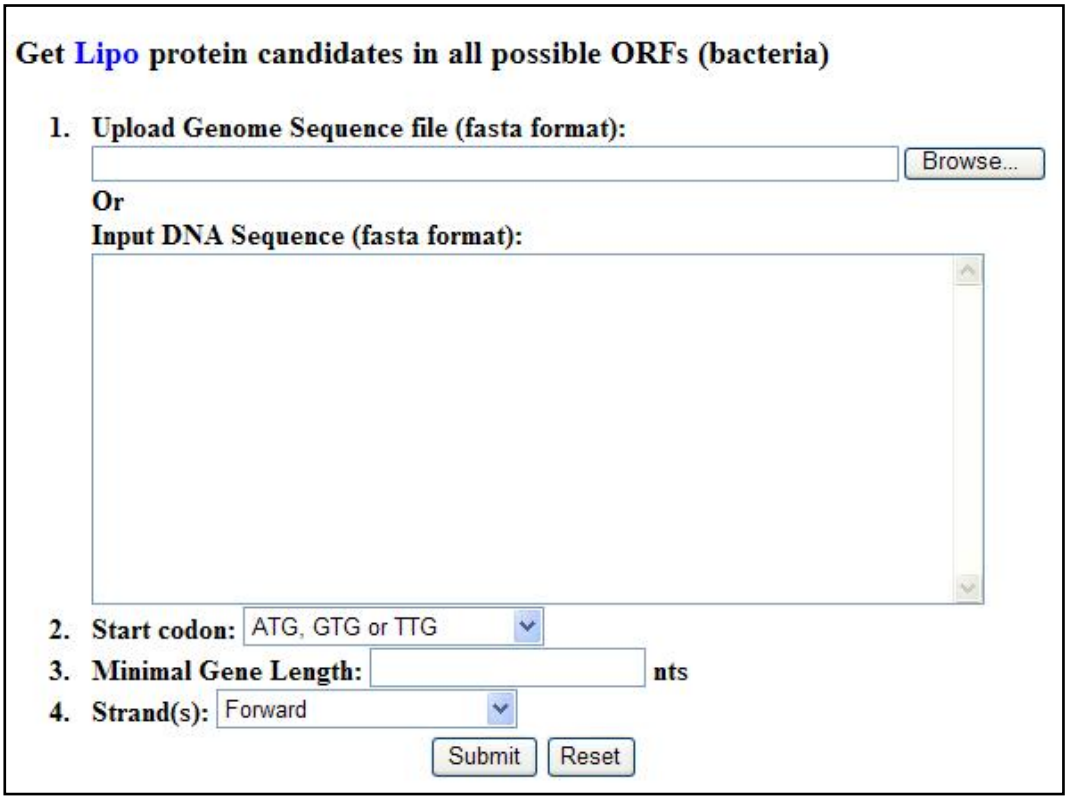


Figure 5.7. The form to submit lipoproteins searching in LipoRy.

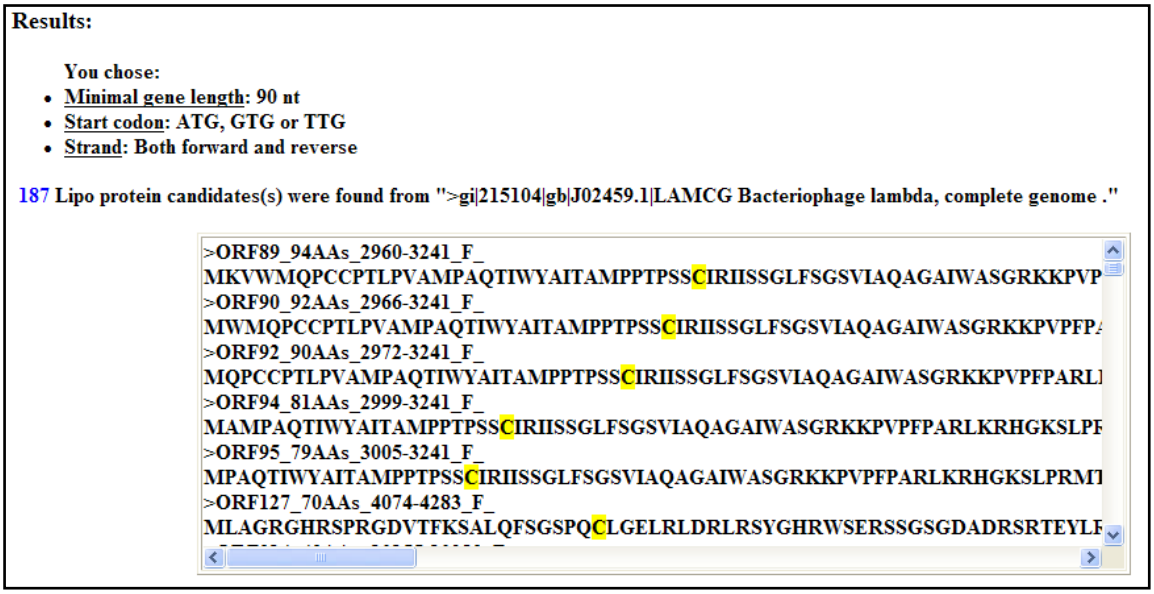


Figure 5.8. Results of lipoprotein searching for the complete genome of bacteriophage lambda. 187 candidates for lipoproteins were found. Cysteines following the signal sequence are highlighted in yellow.

GetRzRz1

Rz and *Rz1* genes are found in most dsDNA and dsRNA bacteriophages. The Rz protein is widely annotated as an endopeptidase, which might cleave the cell wall and the Lpp lipoproteins in the outer membrane of Gram-negative bacteria. *Rz1* gene is located in or closely after *Rz* gene sequence. The Rz1 protein is out membrane lipoprotein. Proteins encoded by both *Rz* and *Rz1* are required for lysis since both *Rz* and *Rz1* mutations block lysis of bacteriophage lambda (Zhang and Young, 1999),.

Rz and *Rz1* have been characterized in only a few bacteriophages genomes(Markov *et al.*, 2004). To predict *Rz*/*Rz1* genes, there was no existing program. A project in the R. Young's laboratory is to find *Rz*/*Rz1* genes in all sequenced phage genomes. Based on the specific character of *Rz* and *Rz1* genes, GetRzRz1 (http://dimer.tamu.edu/cgi-bin/phage_genomics/Rz_Rz1.pl) was written to predict pairs of *Rz* and *Rz1* genes. Prediction of *Rz1* genes is the same as LipoRy, which predicts lipoproteins of Gram-negative bacteria and their phages. To predict *Rz* genes, the following rules were generated by R. Young and E. Summer: 1) there must be a transmembrane domain (TMD) with at least 16 contiguous non-charged amino acid except positions +4 and -4, which could be lysine; 2) there must be only one TMD; 3) the TMD must be within the first 40 amino acids from the N-terminal end; 4) there must be at least one positively charged residue proceeding the TMD; 5) a lysine (K) at the N-terminus of the TMD counts for the positively charged amino acid. The process begins by finding all possible ORFs based on the input DNA sequence. Next, lipoproteins and transmembrane proteins were predicted. Then the *Rz* and *Rz1* gene pairs of were predicted. The maximum

distance between *Rz* and Rz1 genes is set at 200nt. The input of GetRzRz1 is similar to LipoRy (Fig. 5.7). In the output of GetRzRz1, the gene pairs of *Rz* and *Rz1* are listed with information of gene coordinates, length of protein and protein sequence (Fig. 5.9). To test it, the genome sequence enterobacteria phage lambda was searched using this program and 25 *Rz/Rz1* gene pair candidates were found. Out of the 25 pairs, many Rz or Rz1 candidates are proteins with same sequences but different N-terminus, so some of the pairs actually contain the same candidate gene pair. As a result, 13 *Rz/Rz1* gene pair candidates were found, among which was the known *R z/Rz1* gene pair was found in these candidates (Fig. 5.9). This program can be applied to any sequenced genome or DNA sequence to search candidate for *Rz/Rz1* gene pair.

Figure 5.9. Results of *Rz/Rz1* pair searching from enterobacteria phage lambda using GetRzRz1. 25 candidates of *Rz/Rz1* pairs were found. Arrow points the real *Rz/Rz1* gene pair of enterobacteria phage lambda.

SAREndolysin

Recently, the signal arrest-release (SAR) domain was characterized in lysozyme Lyz protein of P1 phage (Xu *et al.*, 2005), which is an ortholog of T4 lysozyme (Bonovich and Young, 1991; Schmidt *et al.*, 1996). Lysozymes of many phages, such as, lambdoid coliphage 21 (R^{21}), coliphages Mu (Xu *et al.*, 2004) and T1 (Roberts *et al.*, 2004), have similar motifs as SAR. Sequences of SAR lysozymes have big diversity in different phages (Xu *et al.*, 2005). It was of interest to determine how wide spread SAR domains are in the lysozemes of other phages. In a SAR domain, there is a transmembrane domain (TMD) in the first 40 amino acids of N-terminal domain. Following the TMD, within 25 residues, there is a lysozyme domain. In the T4 type lysozyme family, there is a glutamic acid (E) followed by an aspartic acid or a cysteine at position +10, and a threonine at position +15 or +16. Based on these characteristic of SAR domain, a program (SAREndolysin) (http://dimer.tamu.edu/cgi-bin/phage_genomics/SarEndolysin.pl) was written to predict candidates of SAR endolysin protein from all possible ORFs of a given DNA sequence (Fig. 5.10). To test this program, the genome sequence of enterobacteria phage P1 was searched. Only one hit, *lyz* gene obtained. As a negative control, the complete genome sequence of Bacteriophage lambda was searched and no SAR endolysin was found. This program can be applied to search for SAR endolysins in any DNA sequence.

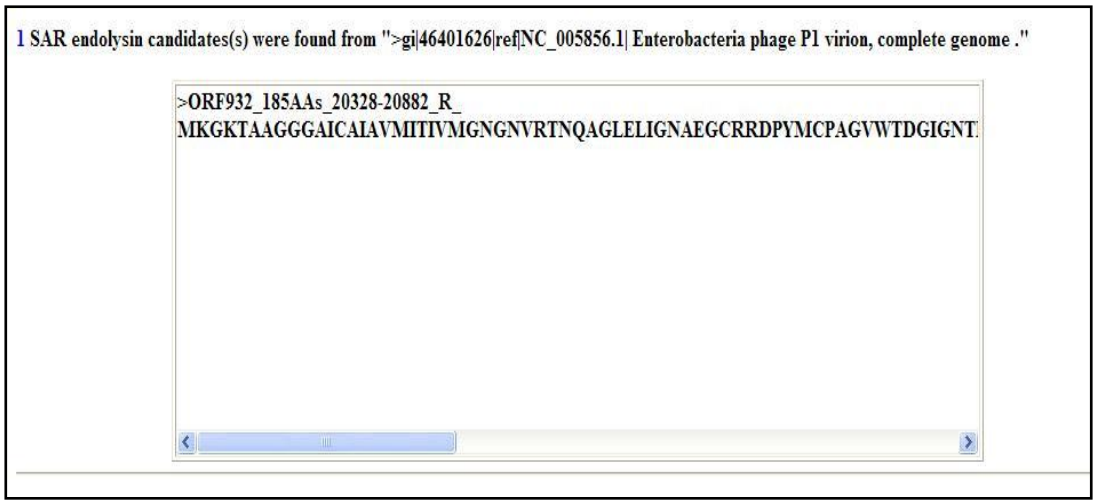


Figure 5.10. The result of SAR endolysin searching from genome of enterobacteria phage P1. One candidate was found which located at 20328 to 20882 in reverse strand.

Primary structure with charges

Current algorithms for predictions of signal sequences and TMD still have a high error rate, so it is imperative to be able to inspect the predicted amino acid sequence manually. Especially important is the ability to see the amino acid sequence and distribution of charged residues. To display the primary structure of amino acid with charges, current existing programs have defined length of output for each line and can only display one protein each time. For comparison of a set of similar proteins, it is more convenient to have a program that could display multiple proteins each time. Moreover to display the output more flexible, which could facilitate comparison between proteins, it is better to have the length of output for each line changeable by users. To satisfy these requirements, the program "primary structure with charges" (http://dimer.tamu.edu/cgi-bin/phage_genomics/charged_aa.pl) was written. After uploading the FASTA protein sequence, the user can choose the number of amino acids per line, and select whether or not to show amino acid position numbers (Fig. 5. 11). The output consists of the primary structure of any ORF, with charge distribution conveniently displayed (Fig. 5. 12).

Amino acid content

Amino acid content is a unique characteristic for proteins, potentially differentiating a specific protein from a group of protein, shared by a set of similar proteins. To calculate amino acid content for both single protein and a set of proteins, a program (http://dimer.tamu.edu/cgi-bin/phage_genomics/AA_content.pl) was written. Charged (in red color), polar (in blue color) and hydrophobic (in yellow color) amino acids are also

grouped (Fig. 5. 13). Moreover, this program can calculate the amino acid content not only for every single protein but also for a set of protein. For example, after calculating and comparing the amino acid content of Rz1 proteins of lambdoid phages and the total proteins of *E. coli*, one could easily find out that Rz1 proteins have higher percentage of cysteine and proline.

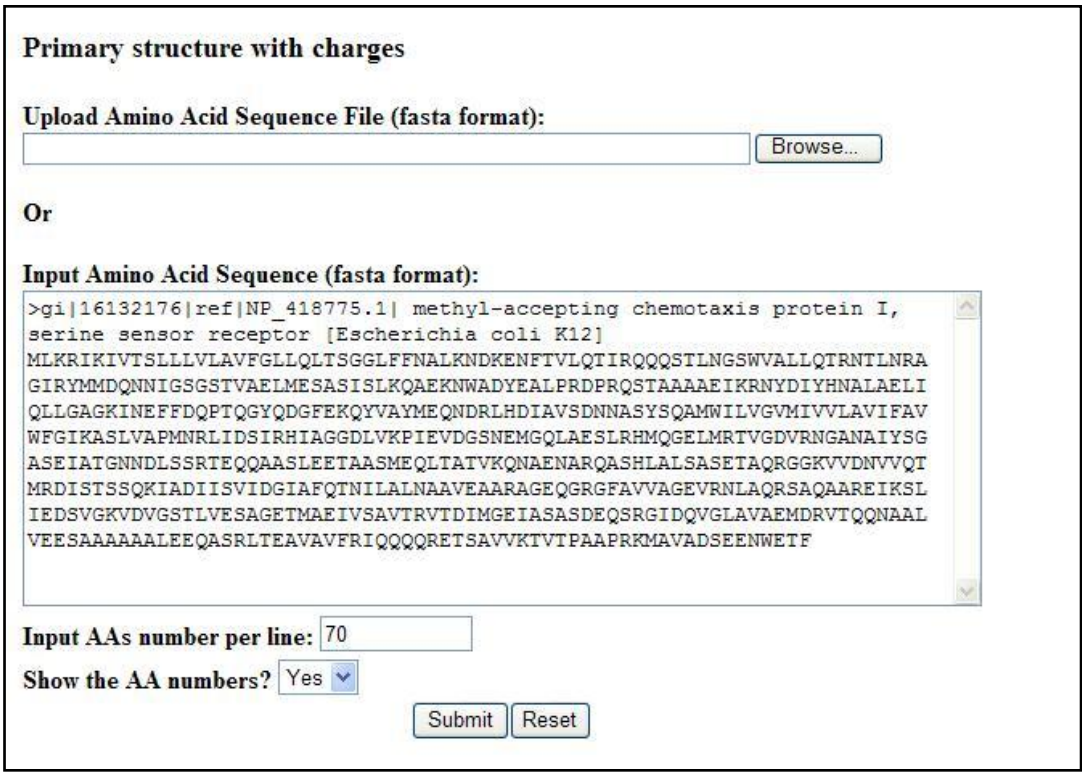


Figure 5.11. The form to display primary structure of protein sequence.

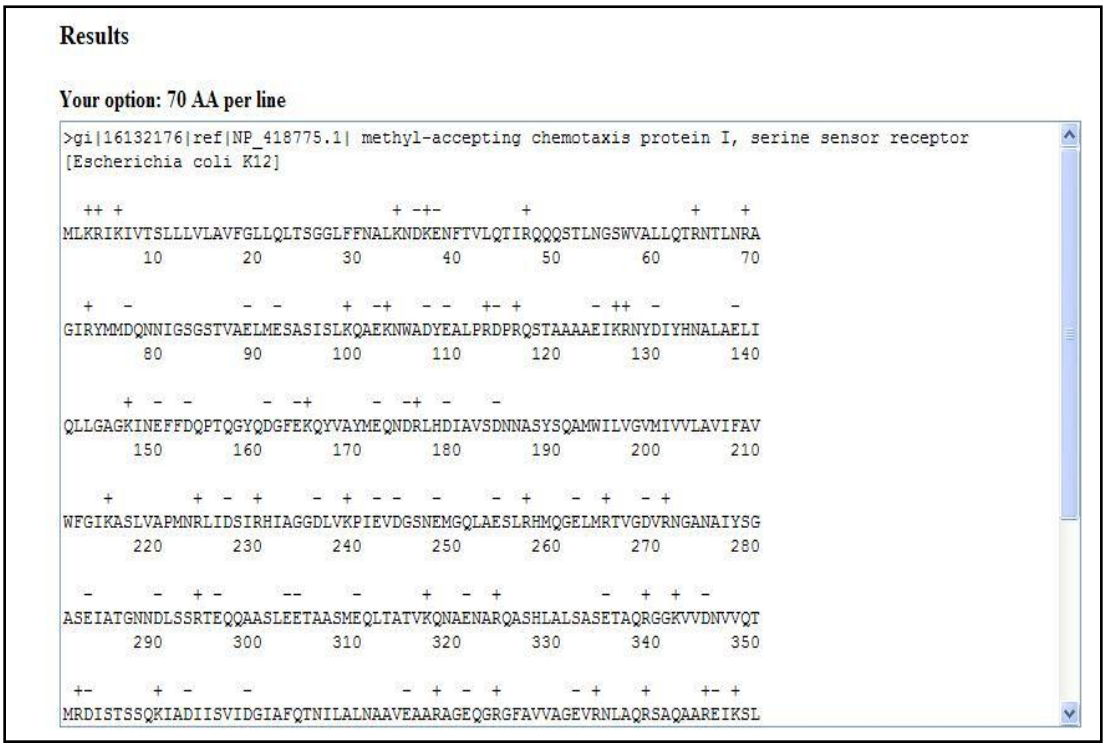
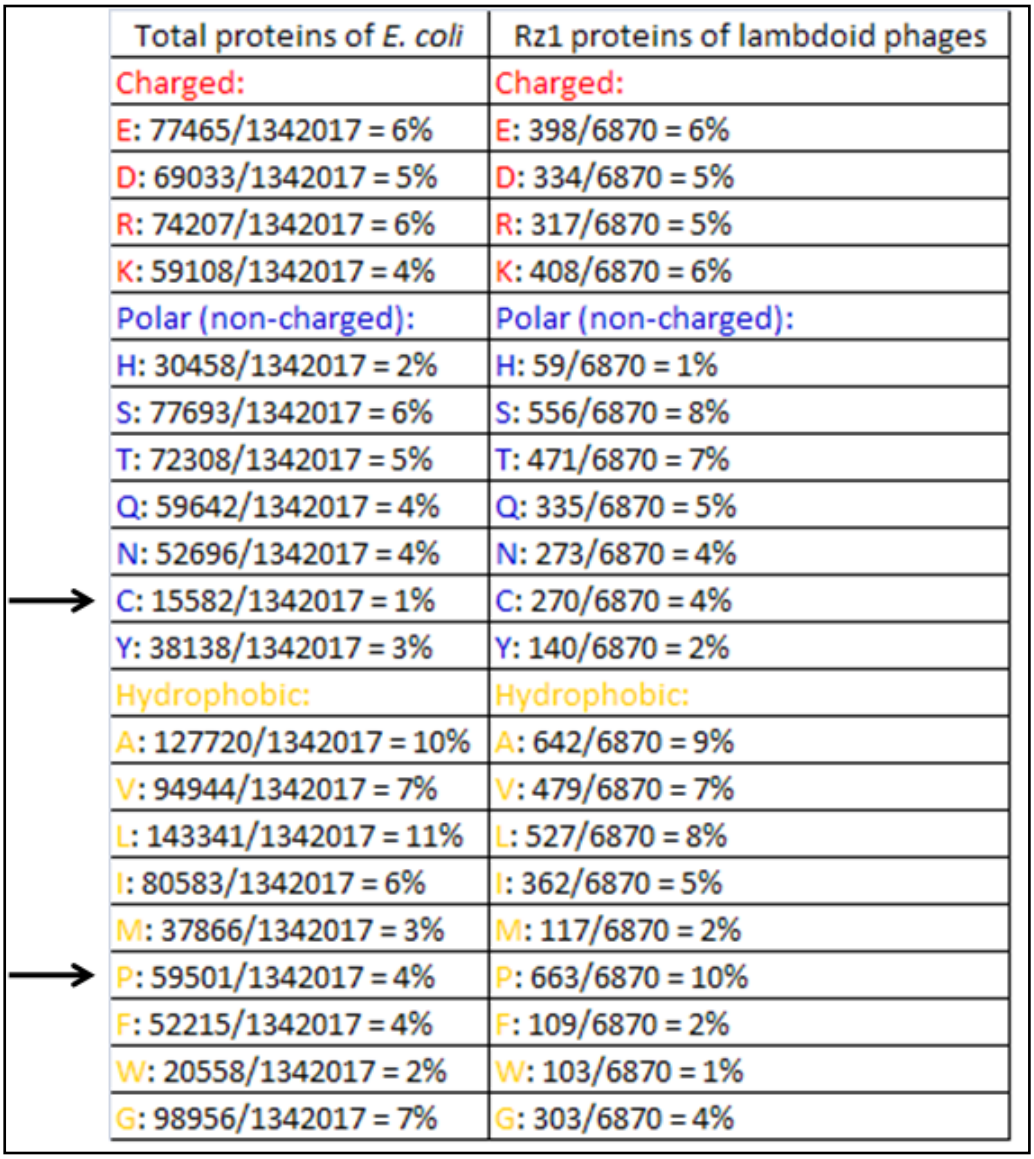


Figure 5.12. Display of primary structure with charges for serine sensor receptor of *E. coli* K-12.



	Total proteins of *E. coli*	Rz1 proteins of lambdoid phages
	Charged:	Charged:
	E: 77465/1342017 = 6%	E: 398/6870 = 6%
	D: 69033/1342017 = 5%	D: 334/6870 = 5%
	R: 74207/1342017 = 6%	R: 317/6870 = 5%
	K: 59108/1342017 = 4%	K: 408/6870 = 6%
	Polar (non-charged):	Polar (non-charged):
	H: 30458/1342017 = 2%	H: 59/6870 = 1%
	S: 77693/1342017 = 6%	S: 556/6870 = 8%
	T: 72308/1342017 = 5%	T: 471/6870 = 7%
	Q: 59642/1342017 = 4%	Q: 335/6870 = 5%
	N: 52696/1342017 = 4%	N: 273/6870 = 4%
→	C: 15582/1342017 = 1%	C: 270/6870 = 4%
	Y: 38138/1342017 = 3%	Y: 140/6870 = 2%
	Hydrophobic:	Hydrophobic:
	A: 127720/1342017 = 10%	A: 642/6870 = 9%
	V: 94944/1342017 = 7%	V: 479/6870 = 7%
	L: 143341/1342017 = 11%	L: 527/6870 = 8%
	I: 80583/1342017 = 6%	I: 362/6870 = 5%
	M: 37866/1342017 = 3%	M: 117/6870 = 2%
→	P: 59501/1342017 = 4%	P: 663/6870 = 10%
	F: 52215/1342017 = 4%	F: 109/6870 = 2%
	W: 20558/1342017 = 2%	W: 103/6870 = 1%
	G: 98956/1342017 = 7%	G: 303/6870 = 4%

Figure 5.13. Calculation of amino acid content of Rz1 proteins of lambdoid phages versus the total proteins of *E. coli*. Charged, polar and hydrophobic amino acids are grouped. Arrows point amino acids, cysteine and proline, with high percentage in Rz1 proteins compared with the total proteins of *E. coli*.

GroupProtein

There are many programs which process multiple protein sequence alignment, such as ClustalW, MULTALIAN (Batzoglou, 2005). To be accurate most of these programs require the sequences are closely related (Higgins *et al.*, 1996). However for phage proteins with similar function, their protein sequences are widely diverse and hard to align accurately with available software. One way to solve this problem is to group these proteins into sub-families that have related sequences, and then apply ClustalW or other program to align each sub-family and draw a family evolution "tree".

To group proteins into sub-families a program (http://dimer.tamu.edu/cgi-bin/phage_genomics/blast_group_aa.pl) was written using BLAST as a tool. BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1990; Pertsemlidis and Fondon, 2001), is a tool to search similarity between nucleotide and/or protein sequences using an algorithm to detect local as well as global alignments. There are many applications of BLAST for genome annotations and functional and structural annotation of proteins (Muller *et al.*, 1999). Functions and even structures (Muller *et al.*, 1999; Song *et al.*, 2006) of unknown proteins could be predicted by BLASTing with nucleotide and/or protein sequences databases.

BLAST can be applied to assess similarity between proteins (Corpet *et al.*, 1998, 1999; Remm and Sonnhammer, 2000). Using the BLAST score of sequence similarities, proteins could be grouped into families. If protein A is similar to protein B, protein B is similar to protein C, then protein A, B and C will be grouped into the same family. To do this work, a program "GroupProtein" was written. The input of this program contains

protein FASTA sequences, name of the query, and E-value for BLAST (Fig. 5.14). Proteins will be compared by BLAST and grouped into families (Fig. 5.15). BLAST result is also retrievable (Fig. 5.15). This program was tested using a mixture of 92 Rz proteins and 108 Rz1 proteins. As a result, Rz proteins were grouped with Rz proteins and Rz1 proteins were group with Rz1 proteins. No Rz protein was grouped with Rz1 proteins.

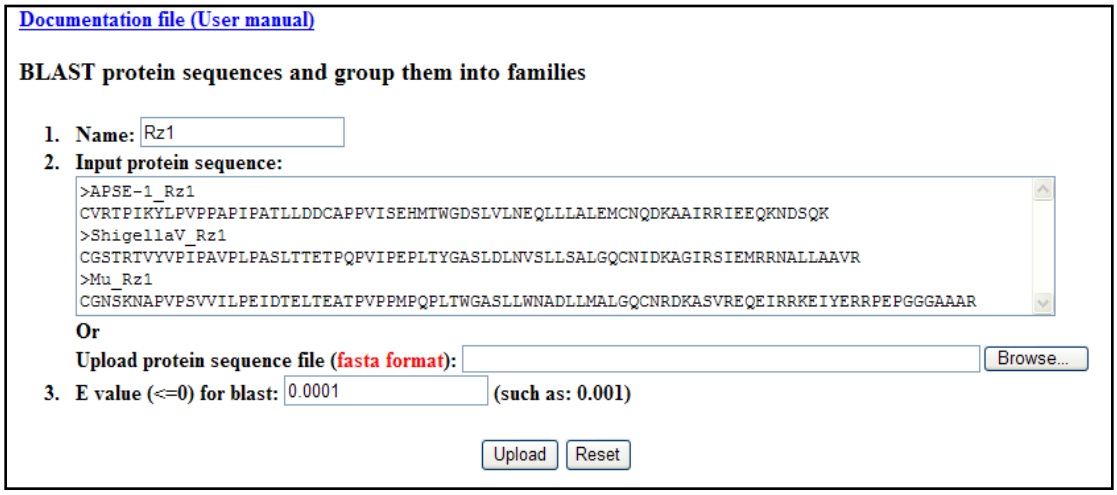


Figure 5.14. The form for "GroupProtein". Protein FASTA sequences of APSE-1 Rz1, ShigellaV Rz1, and Mu Rz 1 were input. E-value was set as 0.0001.

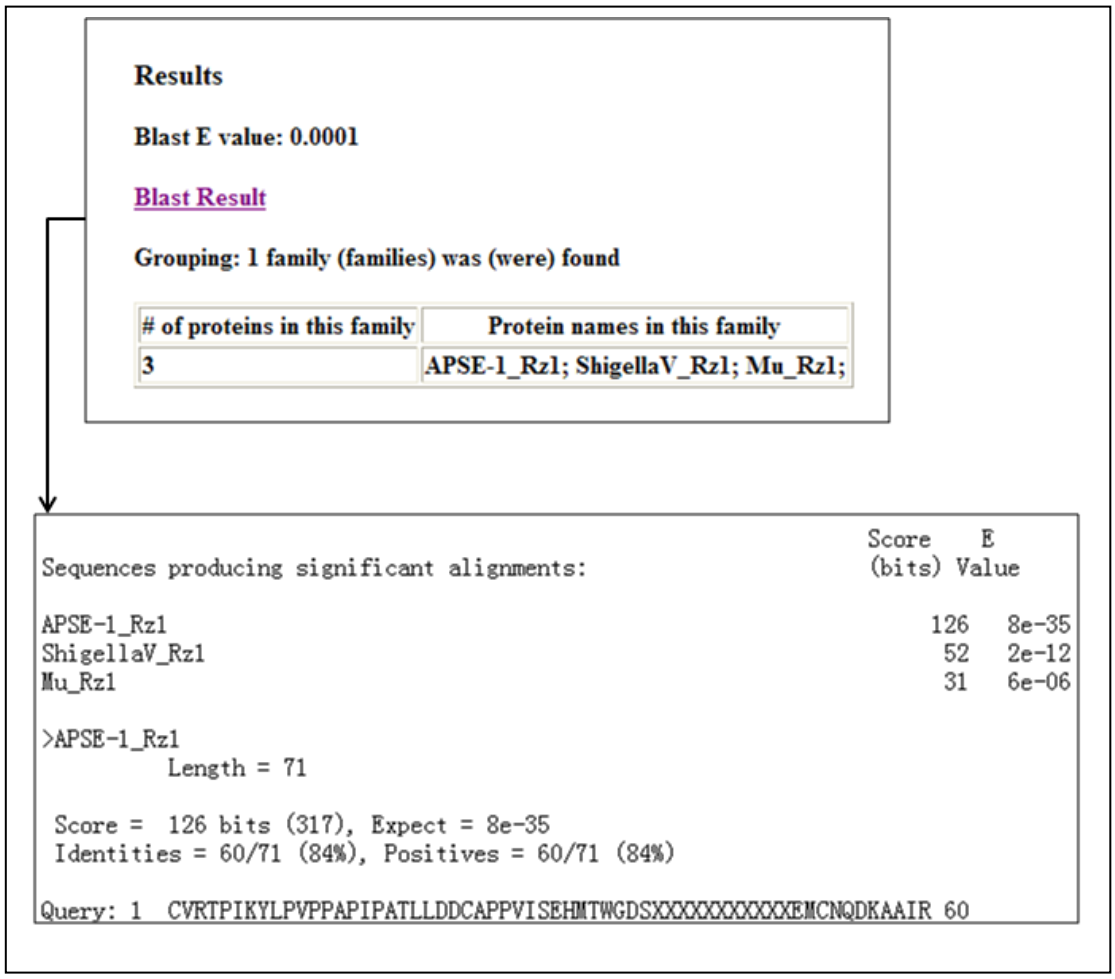
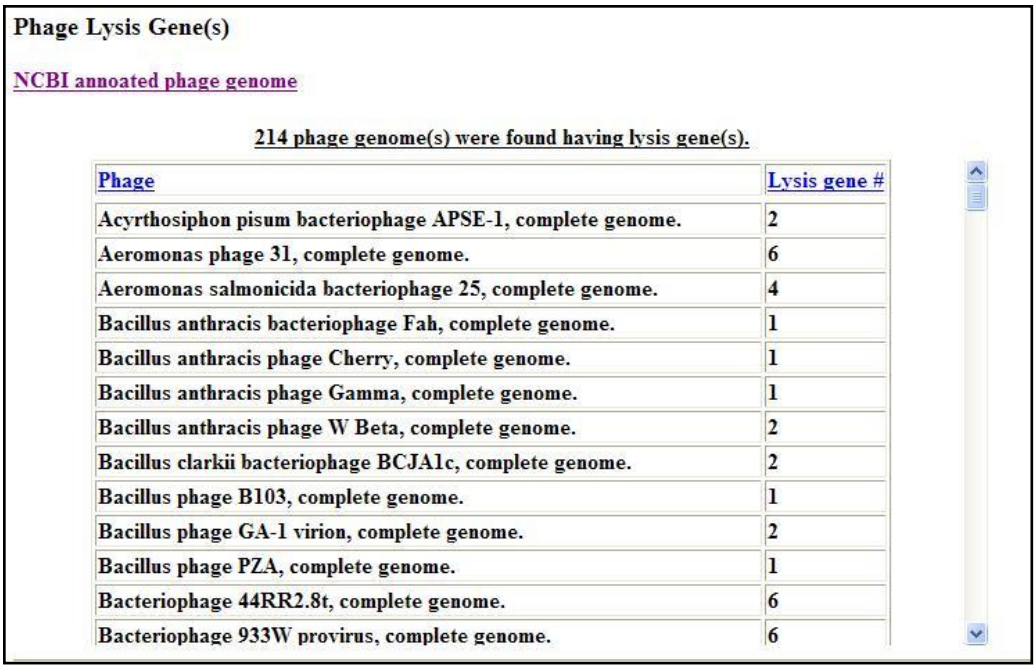


Figure 5.15. Results of BLASTing and grouping APSE-1 Rz1, ShigellaV Rz1, and Mu Rz 1 using "GroupProtein". These three proteins were grouped as one family and the BLAST result is retrievable in this result.

Phage lysis genes

Lysis genes have identified in many phage genomes. The common types of lysis genes are holin, antiholine, endolysin, and Rz/Rz1. To build a web-based database of all the annotated phage lysis genes (http://www.ncbi.nlm.nih.gov/genomes/static/phg.html), programs were written for different steps. First, annotation page for every annotated phage genome was searched. For example, for Acholeplasma phage L2, the annotation page is located at http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi??db=nucleotide&val=NC_001447. Any annotated gene in which the text contains the following word, "lysis", "lysozyme", "lysin", "holin", "antiholine", "endolysin", "Rz", "Rz1", or "endopeptidase", are selected as candidates for phage lysis genes. Second, the results were stored in a MySQL database and displayed on a web page (http://dimer.tamu.edu/cgi-bin/phage_genomics/phage_lysis_genes.pl). In the first part, phage genomes with candidates of phage lysis gene(s) were displayed (Fig. 5.16). By August, 2006, out of 405 annotated phage genomes, 214 of them have one to ten lysis gene candidates, while 191 phage genomes have no annotated lysis gene candidates. In the second part, details of every candidate, such as phage genome name, Genbank Accession ID, version, gene, CDS, locus, product, protein ID, protein GI, note and date, is shown (Fig. 5.17). By August, 2006, 458 candidates for lysis genes were found from 214 annotated phage genomes.



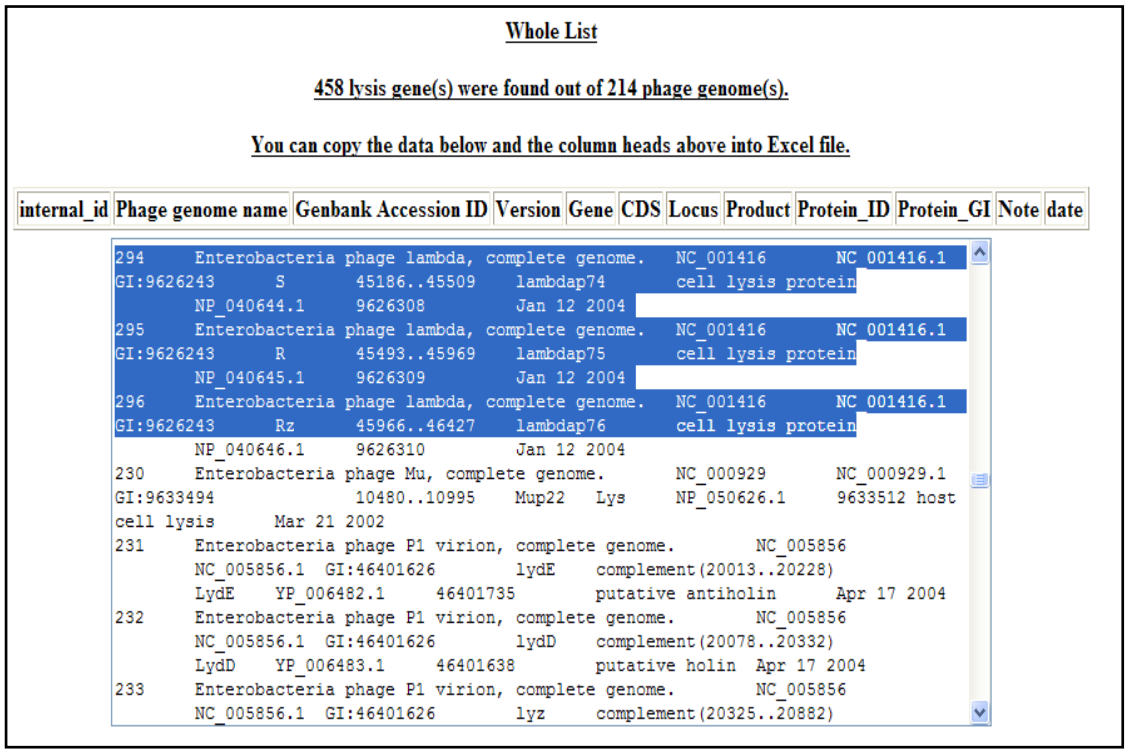
Phage Lysis Gene(s)

NCBI annoated phage genome

214 phage genome(s) were found having lysis gene(s).

Phage	Lysis gene #
Acyrthosiphon pisum bacteriophage APSE-1, complete genome.	2
Aeromonas phage 31, complete genome.	6
Aeromonas salmonicida bacteriophage 25, complete genome.	4
Bacillus anthracis bacteriophage Fah, complete genome.	1
Bacillus anthracis phage Cherry, complete genome.	1
Bacillus anthracis phage Gamma, complete genome.	1
Bacillus anthracis phage W Beta, complete genome.	2
Bacillus clarkii bacteriophage BCJA1c, complete genome.	2
Bacillus phage B103, complete genome.	1
Bacillus phage GA-1 virion, complete genome.	2
Bacillus phage PZA, complete genome.	1
Bacteriophage 44RR2.8t, complete genome.	6
Bacteriophage 933W provirus, complete genome.	6

Figure 5.16. Phage genomes with candidates of lysis gene(s). By Aug. 2006, 214 annotated phage genomes were found having candidates of lysis genes.



Whole List

458 lysis gene(s) were found out of 214 phage genome(s).

You can copy the data below and the column heads above into Excel file.

internal_id	Phage genome name	Genbank Accession ID	Version	Gene	CDS	Locus	Product	Protein_ID	Protein_GI	Note	date

```
294     Enterobacteria phage lambda, complete genome.   NC_001416       NC_001416.1
GI:9626243      S       45186..45509    lambdap74       cell lysis protein
        NP_040644.1     9626308         Jan 12 2004
295     Enterobacteria phage lambda, complete genome.   NC_001416       NC_001416.1
GI:9626243      R       45493..45969    lambdap75       cell lysis protein
        NP_040645.1     9626309         Jan 12 2004
296     Enterobacteria phage lambda, complete genome.   NC_001416       NC_001416.1
GI:9626243      Rz      45966..46427    lambdap76       cell lysis protein
        NP_040646.1     9626310         Jan 12 2004
230     Enterobacteria phage Mu, complete genome.       NC_000929       NC_000929.1
GI:9633494              10480..10995    Mup22   Lys     NP_050626.1     9633512 host
cell lysis      Mar 21 2002
231     Enterobacteria phage P1 virion, complete genome.        NC_005856
        NC_005856.1  GI:46401626        lydE    complement(20013..20228)
        LydE    YP_006482.1     46401735        putative antiholin      Apr 17 2004
232     Enterobacteria phage P1 virion, complete genome.        NC_005856
        NC_005856.1  GI:46401626        lydD    complement(20078..20332)
        LydD    YP_006483.1     46401638        putative holin  Apr 17 2004
233     Enterobacteria phage P1 virion, complete genome.        NC_005856
        NC_005856.1  GI:46401626        lyz     complement(20325..20882)
```

Figure 5.17. Details of candidate phage lysis genes from annotated phage genomes. Candidates highlighted are the three lysis genes, S, R and Rz, found for enterobacteria phage lambda,

Discussion

For specific requirements in the annotation of the genome of *E. coli* and a number of bacteriophages, I generated broad array of programs. Using Gbrowse, the genome of *E. coli* MG1655, *E. coli* W3110, and 28 bacteriophages were displayed. Programs were also generated for prediction of lipoproteins, *Rz* and *Rz1* genes, and SAR endolysins.

LipoRy and GetRzRz1 have relative high false positive rates. One reason is the first step of these programs searches all possible ORFs, instead of genes. However, some phage genomes are just recently sequenced and have crude or no-existed annotation. Without knowing the properties of genes, it is hard to add restrictions to filter non-gene ORFs. For the first step of searching specific genes, false positive is better than false negative. Before GetORF was generated, programs have been written to search for Shine-Dalgarno (S-D) motif sequences in several phage genomes. There are some common S-D motifs such as AGGAG, but the S-D motifs of large proportion genes are very diverse. With more and more knowledge of properties of phage genes and specific genes, programs such as LipoRy and GetRzRz1 can be improved with less false positive with better filters.

GroupProtein is a program to group proteins into families using the results from BLAST. The motivation to write this program is from the large diversity of phage protein sequences even for those within protein families sharing the same function. To be useful, programs like ClustalW require sequences to be closely related. To align proteins with lower sequence similarity, such as phage proteins, GroupProtein can be applied to group proteins into sub-families first. For each sub-family, which has relative high

similarity in sequence, other alignment programs can be applied.

Rz/*Rz1*, and SAR endolysin gene are all phage lysis genes. Besides programs written to predict these genes, it is also critical to see how many lysis genes are already annotated in the current phage genomes. Programs were written to search all the candidates for phage lysis genes in current annotated phage genomes. Out of 405 annotated phage genomes, 301 of them have only one or no annotated lysis gene at all. Based on this result, there are many lysis genes need to be annotated even in annotated phage genomes. Summary of the current lysis genes in all annotated phage genomes can help scientist focus on genomes that need more effort to improve the annotation.

Although genome annotation of bacteria and phages cannot be done automatically by any current computer program, with more knowledge of specific protein properties, more programs could be generated to make this process easier and faster. With proper combination of different programs, the process of genome annotation could be mad semi-automatic or, eventually, automatic.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

To study changes happen when *E. coli* cells enter stationary phase, two different subproteome of *E. coli* cells were analyzed using heparin and AIX columns. Also bioinformatics tools were utilized to facilitate our proteomic studies. In total 318 heparin-binding proteins and 61 AIX-binding proteins were identified. 116 heparin-binding proteins and 2 AIX-binding proteins were newly identified compared with other proteomics studies (Champion *et al.*, 2003; Corbin *et al.*, 2003; Sundararaj *et al.*, 2004; Tonella *et al.*, 2001). Heparin enriched and identified low abundance proteins. The AIX/2D PAGE study provided a novel way to study protein abundance and chromatography behavior. Thousands of proteins spots on 38 2D PAGE of continuous AIX fractions were visualized for both exponential and stationary phase samples. Elution profile and abundance for same spot on different 2D PAGE were obtained using Z3 gel image software and Perl script.

Using heparin columns and AIX columns, differences were found in protein presence, protein abundance and chromatography behavior between exponential and stationary phase of *E. coli* cells. Changes also happened to proteins in different functional groups. For protein presence, more proteins were found in functional groups of carbohydrate transport and metabolism, post-translational modification, protein turnover, chaperones, and inorganic ion transport and metabolism for stationary phase

Heparome. Five heparin-binding proteins and twenty-five AIX-binding proteins were found more abundant in stationary phase. These proteins with more abundance in stationary phase are mainly in functional groups of chaperones, starvation stresses, amino acid transport and metabolism, carbohydrate transport and metabolism and translation. Also seven proteins were found with different elution profiles on heparin columns between exponential and stationary phase.

Web-based databases, EEP, were generated to manage and mine our data more accurately and efficiently. Using the website, our data is possible to be shared with others. EEP is the first *E. coli* proteomic database, which include data from different E. *coli* proteomic studies. By including data from other groups into our databases, comparisons of different proteomic studies were made.

Discussion and future prospects

Heparome and AIX/2D PAGE identified protein with different amount and different elution positions on columns for samples from exponential growth and stationary phase. More work should be done to study the reasons that cause the difference, which is related to the physiology of *E. coli* cells in stationary phase. For example, FtnA were found at different elution positions from heparin column between these two growth phases. Elution positions of FtnA were also verified by western blot. Reason causing the elution shift might be protein conformation, post-translational modification or protein-protein interactions.

Heparome and AIX/2D PAGE increased the coverage of whole *E. coli* proteome by 118 newly identified proteins. However the predicted proteome of *E. coli* has more than

4000 gene products. Although not all of them expressed under all conditions, global proteomics are biased to high abundant proteins. To increase the coverage of *E. coli* proteome, more growth conditions and subproteomics should be applied.

2D PAGE has advantages to study amount, physical properties (Obs. pI and Obs. MW) and post-translational modifications of proteins. 18 proteins were found in multiple protein spots on gel by AIX/2D PAGE. 6 of these proteins, AceA, AhpC, FusA, RpsA, Tig, TufA, have multiple proteins spots with different abundant trend between exponential and stationary phase. Isoform of protein on 2D PAGE could be caused by protein degradation or post-translational modifications.

Linkage between multiple 2D PAGE using Z2 2D gel analysis system and computational language is efficient. Although there are still technique difficulties to study proteins in multiple fractions and 2D PAGE, new technologies, such as CyDye DIGE 2D PAGE and robotic gel visualization and spot cutting could be applied to improve this study.

The web-based database is a modern way to manage large amount of data generated by "OMICS" studies. As more and more proteomics studies will be accomplished, proteomic databases, such as EEP, which will handle data merging, would be important to trace the current status of proteomics study. This will also help scientists to find out what is lack in proteomics study.

REFERENCES

Adams, M.A., Iannuzzi, P., and Jia, Z. (2005) MdaB from *Escherichia coli*: cloning, purification, crystallization and preliminary X-ray analysis. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **61**: 235-238.

Adams, M.A., and Jia, Z. (2006) Modulator of drug activity B from *Escherichia coli*: crystal structure of a prokaryotic homologue of DT-diaphorase. *J Mol Biol* **359**: 455-465.

Aebersold, R., and Goodlett, D.R. (2001) Mass spectrometry in proteomics. *Chem Rev* **101**: 269-295.

Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**: 198-207.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Anderson, L., and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**: 533-537.

Anderson, N.L., and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**: 1853-1861.

Andrews, S.C. (1998) Iron storage in bacteria. *Adv Microb Physiol* **40**: 281-351.

Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H.C., Hirai, A., Tsuzuki, K., Nakamura, S., Altaf-Ul-Amin, M., Oshima, T., Baba, T., Yamamoto, N., Kawamura, T., Ioka-Nakamichi, T., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006) Large-scale

identification of protein-protein interaction of *Escherichia coli* K-12. *Genome Res* **16**: 686-691.

Armstrong, D.J., and Roman, A. (1992) Mutagenesis of human papillomavirus types 6 and 16 E7 open reading frames alters the electrophoretic mobility of the expressed proteins. *J Gen Virol* **73 (Pt 5)**: 1275-1279.

Arnold, C.N., McElhanon, J., Lee, A., Leonhart, R., and Siegele, D.A. (2001) Global analysis of *Escherichia coli* gene expression during the acetate-induced acid tolerance response. *J Bacteriol* **183**: 2178-2186.

Avetisyan, A.V., Dibrov, P.A., Semeykina, A.L., Skulachev, V.P., and Sokolov, M.V. (1991) Adaptation of Bacillus FTU and *Escherichia coli* to alkaline conditions: the Na(+)-motive respiration. *Biochim Biophys Acta* **1098**: 95-104.

Aviv, M., Giladi, H., Oppenheim, A.B., and Glaser, G. (1996) Analysis of the shut-off of ribosomal RNA promoters in *Escherichia coli* upon entering the stationary phase of growth. *FEMS Microbiol Lett* **140**: 71-76.

Baclawski, K., Futrelle, R., Fridman, N., and Pescitelli, M.J. (1993) Database techniques for biological materials & methods. *Proc Int Conf Intell Syst Mol Biol* **1**: 21-28.

Badock, V., Steinhusen, U., Bommert, K., and Otto, A. (2001) Prefractionation of protein samples for proteome analysis using reversed-phase high-performance liquid chromatography. *Electrophoresis* **22**: 2856-2864.

Baglioni, P., Bini, L., Liberatori, S., Pallini, V., and Marri, L. (2003) Proteome analysis of *Escherichia coli* W3110 expressing an heterologous sigma factor. *Proteomics* **3**: 1060-1065.

Ballesteros, M., Fredriksson, A., Henriksson, J., and Nystrom, T. (2001) Bacterial senescence: protein oxidation in non-proliferating cells is dictated by the accuracy of the ribosomes. *Embo J* **20**: 5280-5289.

Batzoglou, S. (2005) The many faces of sequence alignment. *Brief Bioinform* **6**: 6-22.

Baumgart, S., Lindner, Y., Kuhne, R., Oberemm, A., Wenschuh, H., and Krause, E. (2004) The contributions of specific amino acid side chains to signal intensities of peptides in matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* **18**: 863-868.

Berven, F.S., Karlsen, O.A., Straume, A.H., Flikka, K., Murrell, J.C., Fjellbirkeland, A., Lillehaug, J.R., Eidhammer, I., and Jensen, H.B. (2006) Analysing the outer membrane subproteome of Methylococcus capsulatus (Bath) using proteomics and novel biocomputing tools. *Arch Microbiol* **184**: 362-377.

Bird, J.G., Sharma, S., Roshwalb, S.C., Hoskins, J.R., and Wickner, S. (2006) Functional Analysis of CbpA, a DnaJ Homolog and Nucleoid-associated DNA-binding Protein. *J Biol Chem* **281**: 34349-34356.

Bonovich, M.T., and Young, R. (1991) Dual start motif in two lambdoid S genes unrelated to lambda S. *J Bacteriol* **173**: 2897-2905.

Borodovsky, M. (1998) 'Small genomics' on a large scale. E. coli and small genomes, American Society for Microbiology, Snowbird, UT, USA, 12-15 October 1997. *Trends Genet* **14**: 51-52.

Bunai, K., and Yamane, K. (2005) Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives. *J*

Chromatogr B Analyt Technol Biomed Life Sci **815**: 227-236.

Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**: 531-537.

Cao, G.J., and Sarkar, N. (1997) Stationary phase-specific mRNAs in *Escherichia coli* are polyadenylated. *Biochem Biophys Res Commun* **239**: 46-50.

Chae, C., Sharma, S., Hoskins, J.R., and Wickner, S. (2004) CbpA, a DnaJ homolog, is a DnaK co-chaperone, and its activity is modulated by CbpM. *J Biol Chem* **279**: 33147-33153.

Champion, M.M., Campbell, C.S., Siegele, D.A., Russell, D.H., and Hu, J.C. (2003) Proteome analysis of *Escherichia coli* K-12 by two-dimensional native-state chromatography and MALDI-MS. *Mol Microbiol* **47**: 383-396.

Charoenlap, N., Eiamphungporn, W., Chauvatcharin, N., Utamapongchai, S., Vattanaviboon, P., and Mongkolsuk, S. (2005) OxyR mediated compensatory expression between ahpC and katA and the significance of ahpC in protection from hydrogen peroxide in Xanthomonas campestris. *FEMS Microbiol Lett* **249**: 73-78.

Chaudhuri, R.R., Khan, A.M., and Pallen, M.J. (2004) coliBASE: an online database for *Escherichia coli*, Shigella and Salmonella comparative genomics. *Nucleic Acids Res* **32**: D296-299.

Chen, L., Xie, Q.W., and Nathan, C. (1998) Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. *Mol Cell* **1**:

795-805.

Cohen, S.L., and Chait, B.T. (1996) Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal Chem* **68**: 31-37.

Corbin, R.W., Paliy, O., Yang, F., Shabanowitz, J., Platt, M., Lyons, C.E., Jr., Root, K., McAuliffe, J., Jordan, M.I., Kustu, S., Soupene, E., and Hunt, D.F. (2003) Toward a protein profile of *Escherichia coli*: comparison to its transcription profile. *Proc Natl Acad Sci U S A* **100**: 9232-9237.

Cordwell, S.J., Nouwens, A.S., Verrills, N.M., Basseal, D.J., and Walsh, B.J. (2000) Subproteomics based upon protein cellular location and relative solubilities in conjunction with composite two-dimensional electrophoresis gels. *Electrophoresis* **21**: 1094-1103.

Corpet, F., Gouzy, J., and Kahn, D. (1998) The ProDom database of protein domain families. *Nucleic Acids Res* **26**: 323-326.

Corpet, F., Gouzy, J., and Kahn, D. (1999) Recent improvements of the ProDom database of protein domain families. *Nucleic Acids Res* **27**: 263-267.

Cottrell, J.S. (1994) Protein identification by peptide mass fingerprinting. *Pept Res* **7**: 115-124.

Cullum, A.J., Bennett, A.F., and Lenski, R.E. (2001) Evolutionary adaptation to temperature. IX. Preadaptation to novel stressful environments of *Escherichia coli* adapted to high temperature. *Evolution Int J Org Evolution* **55**: 2194-2202.

Daley, D.O., Rapp, M., Granseth, E., Melen, K., Drew, D., and von Heijne, G. (2005) Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308**:

1321-1323.

Death, A., and Ferenci, T. (1994) Between feast and famine: endogenous inducer synthesis in the adaptation of *Escherichia coli* to growth with limiting carbohydrates. *J Bacteriol* **176**: 5101-5107.

Deckers, D., Masschalck, B., Aertsen, A., Callewaert, L., Van Tiggelen, C.G., Atanassova, M., and Michiels, C.W. (2004) Periplasmic lysozyme inhibitor contributes to lysozyme resistance in *Escherichia coli*. *Cell Mol Life Sci* **61**: 1229-1237.

Demirev, P.A., Lin, J.S., Pineda, F.J., and Fenselaut, C. (2001) Bioinformatics and mass spectrometry for microorganism identification: proteome-wide post-translational modifications and database search algorithms for characterization of intact H. pylori. *Anal Chem* **73**: 4566-4573.

Desnues, B., Cuny, C., Gregori, G., Dukan, S., Aguilaniu, H., and Nystrom, T. (2003) Differential oxidative damage and expression of stress defence regulons in culturable and non-culturable *Escherichia coli* cells. *EMBO Rep* **4**: 400-404.

Dobrindt, U. (2005) (Patho-)Genomics of *Escherichia coli*. *Int J Med Microbiol* **295**: 357-371.

Duan, J., Liang, Z., Yang, C., Zhang, J., Zhang, L., Zhang, W., and Zhang, Y. (2006) Rapid protein identification using monolithic enzymatic microreactor and LC-ESI-MS/MS. *Proteomics* **6**: 412-419.

Dukan, S., and Touati, D. (1996) Hypochlorous acid stress in *Escherichia coli*: resistance, DNA damage, and comparison with hydrogen peroxide stress. *J Bacteriol* **178**: 6145-6150.

Dukan, S., and Nystrom, T. (1998) Bacterial senescence: stasis results in increased and differential oxidation of cytoplasmic proteins leading to developmental induction of the heat shock regulon. *Genes Dev* **12**: 3431-3441.

Eichel, J., Chang, Y.Y., Riesenberg, D., and Cronan, J.E., Jr. (1999) Effect of ppGpp on *Escherichia coli* cyclopropane fatty acid synthesis is mediated through the RpoS sigma factor (sigmaS). *J Bacteriol* **181**: 572-576.

Eisenstark, A., Calcutt, M.J., Becker-Hapak, M., and Ivanova, A. (1996) Role of *Escherichia coli* rpoS and associated genes in defense against oxidative damage. *Free Radic Biol Med* **21**: 975-993.

Erni, B., and Zanolari, B. (1985) The mannose-permease of the bacterial phosphotransferase system. Gene cloning and purification of the enzyme IIMan/IIIMan complex of *Escherichia coli*. *J Biol Chem* **260**: 15495-15503.

Erni, B., Zanolari, B., and Kocher, H.P. (1987) The mannose permease of *Escherichia coli* consists of three different proteins. Amino acid sequence and function in sugar transport, sugar phosphorylation, and penetration of phage lambda DNA. *J Biol Chem* **262**: 5238-5247.

Eyre-Walker, A. (1996) Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? *Mol Biol Evol* **13**: 864-872.

Ferguson, P.L., and Smith, R.D. (2003) Proteome analysis by mass spectrometry. *Annu Rev Biophys Biomol Struct* **32**: 399-424.

Fountoulakis, M., and Takacs, B. (1998) Design of protein purification pathways: application to the proteome of Haemophilus influenzae using heparin chromatography. *Protein*

Expr Purif **14**: 113-119.

Fredriksson, A., Ballesteros, M., Dukan, S., and Nystrom, T. (2005) Defense against protein carbonylation by DnaK/DnaJ and proteases of the heat shock regulon. *J Bacteriol* **187**: 4207-4213.

Gentry, D.R., Hernandez, V.J., Nguyen, L.H., Jensen, D.B., and Cashel, M. (1993) Synthesis of the stationary-phase sigma factor sigma s is positively regulated by ppGpp. *J Bacteriol* **175**: 7982-7989.

Glasner, J.D., Liss, P., Plunkett, G., 3rd, Darling, A., Prasad, T., Rusch, M., Byrnes, A., Gilson, M., Biehl, B., Blattner, F.R., and Perna, N.T. (2003) ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res* **31**: 147-151.

Graber, A., Juhasz, P.S., Khainovski, N., Parker, K.C., Patterson, D.H., and Martin, S.A. (2004) Result-driven strategies for protein identification and quantitation--a way to optimize experimental design and derive reliable results. *Proteomics* **4**: 474-489.

Gragerov, A., Nudler, E., Komissarova, N., Gaitanaris, G.A., Gottesman, M.E., and Nikiforov, V. (1992) Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in *Escherichia coli*. *Proc Natl Acad Sci U S A* **89**: 10341-10344.

Graham, R.L., O'Loughlin S, N., Pollock, C.E., Ternan, N.G., Weatherly, D.B., Jackson, P.J., Tarleton, R.L., and McMullan, G. (2006) A Combined Shotgun and Multidimensional Proteomic Analysis of the Insoluble Subproteome of the Obligate Thermophile, Geobacillusthermoleovorans T80. *J Proteome Res* **5**: 2465-2473.

Griffin, T.J., Goodlett, D.R., and Aebersold, R. (2001) Advances in proteome analysis by mass spectrometry. *Curr Opin Biotechnol* **12**: 607-612.

Groat, R.G., Schultz, J.E., Zychlinsky, E., Bockman, A., and Matin, A. (1986) Starvation proteins in *Escherichia coli*: kinetics of synthesis and role in starvation survival. *J Bacteriol* **168**: 486-493.

Harris, T.W., Lee, R., Schwarz, E., Bradnam, K., Lawson, D., Chen, W., Blasier, D., Kenny, E., Cunningham, F., Kishore, R., Chan, J., Muller, H.M., Petcherski, A., Thorisson, G., Day, A., Bieri, T., Rogers, A., Chen, C.K., Spieth, J., Sternberg, P., Durbin, R., and Stein, L.D. (2003) WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res* **31**: 133-137.

Hatzimanikatis, V., Choe, L.H., and Lee, K.H. (1999) Proteomics: theoretical and experimental considerations. *Biotechnol Prog* **15**: 312-318.

Hauer, C.R., Staudenmann, W., Kuster, T., Neuheiser, F., Hughes, G.J., Seto-Ohshima, A., Tanokura, M., and Heizmann, C.W. (1992) Protein sequence determination by ESI-MS and LSI-MS tandem mass spectrometry: parvalbumin primary structures from cat, gerbil and monkey skeletal muscle. *Biochim Biophys Acta* **1160**: 1-7.

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B.L., Mori, H., and Horiuchi, T. (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* **2**: 2006 0007.

Hayashi, S., and Wu, H.C. (1990) Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**: 451-471.

Hengge-Aronis, R. (1993) Survival of hunger and stress: the role of rpoS in early stationary phase gene regulation in E. coli. *Cell* **72**: 165-168.

Hengge-Aronis, R. (1996) Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in *Escherichia coli*. *Mol Microbiol* **21**: 887-893.

Hengge-Aronis, R. (2002) Stationary phase gene regulation: what makes an *Escherichia coli* promoter sigmaS-selective? *Curr Opin Microbiol* **5**: 591-595.

Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* **90**: 5011-5015.

Herbert, B.R., Molloy, M.P., Gooley, A.A., Walsh, B.J., Bryson, W.G., and Williams, K.L. (1998) Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent. *Electrophoresis* **19**: 845-851.

Higgins, D.G., Thompson, J.D., and Gibson, T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* **266**: 383-402.

Hightower, L.E. (1991) Heat shock, stress proteins, chaperones, and proteotoxicity. *Cell* **66**: 191-197.

Hsin-Yi, C., and Chou, C.C. (2001) Acid adaptation and temperature effect on the survival of E. coli O157:H7 in acidic fruit juice and lactic fermented milk product. *Int J Food Microbiol* **70**: 189-195.

Hua, Q., Yang, C., Oshima, T., Mori, H., and Shimizu, K. (2004) Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat

cultures. *Appl Environ Microbiol* **70**: 2354-2366.

Huerta, A.M., Salgado, H., Thieffry, D., and Collado-Vides, J. (1998) RegulonDB: a database on transcriptional regulation in *Escherichia coli*. *Nucleic Acids Res* **26**: 55-59.

Ihling, C., and Sinz, A. (2005) Proteome analysis of *Escherichia coli* using high-performance liquid chromatography and Fourier transform ion cyclotron resonance mass spectrometry. *Proteomics* **5**: 2029-2042.

Imai, K., Iida, T., Takano, Y., and Uozumi, N. (2002) Membrane-bound heparin binding proteins from HL-60 cells purified in a two-step affinity chromatography differentially eluted with divalent cations. *J Chromatogr B Analyt Technol Biomed Life Sci* **780**: 1-12.

Ishihama, A. (1997) Adaptation of gene expression in stationary phase bacteria. *Curr Opin Genet Dev* **7**: 582-588.

Iuchi, S., and Weiner, L. (1996) Cellular and molecular physiology of *Escherichia coli* in the adaptation to aerobic environments. *J Biochem (Tokyo)* **120**: 1055-1063.

Izutsu, K., Wada, A., and Wada, C. (2001a) Expression of ribosome modulation factor (RMF) in *Escherichia coli* requires ppGpp. *Genes Cells* **6**: 665-676.

Izutsu, K., Wada, C., Komine, Y., Sako, T., Ueguchi, C., Nakura, S., and Wada, A. (2001b) *Escherichia coli* ribosome-associated protein SRA, whose copy number increases during stationary phase. *J Bacteriol* **183**: 2765-2773.

Juncker, A.S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12**:

1652-1662.

Jung, I.L., and Kim, I.G. (2003) Transcription of ahpC, katG, and katE genes in *Escherichia coli* is regulated by polyamines: polyamine-deficient mutant sensitive to H2O2-induced oxidative damage. *Biochem Biophys Res Commun* **301**: 915-922.

Jung, I.L., Phyo, K.H., and Kim, I.G. (2005) RpoS-mediated growth-dependent expression of the *Escherichia coli* tkt genes encoding transketolases isoenzymes. *Curr Microbiol* **50**: 314-318.

Kabir, M.S., Sagara, T., Oshima, T., Kawagoe, Y., Mori, H., Tsunedomi, R., and Yamada, M. (2004) Effects of mutations in the rpoS gene on cell viability and global gene expression under nitrogen starvation in *Escherichia coli*. *Microbiology* **150**: 2543-2553.

Kahn, P. (1995) From genome to proteome: looking at a cell's proteins. *Science* **270**: 369-370.

Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**: 2299-2301.

Karp, P.D., Riley, M., Paley, S.M., and Pelligrini-Toole, A. (1996) EcoCyc: an encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res* **24**: 32-39.

Karp, P.D., Riley, M., Paley, S.M., Pellegrini-Toole, A., and Krummenacker, M. (1997) EcoCyc: Enyclopedia of *Escherichia coli* Genes and Metabolism. *Nucleic Acids Res* **25**: 43-51.

Karp, P.D., Riley, M., Paley, S.M., Pellegrini-Toole, A., and Krummenacker, M. (1998) EcoCyc: Encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res* **26**: 50-53.

Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Paley, S.M., and Pellegrini-Toole, A. (2000) The EcoCyc and MetaCyc databases. *Nucleic Acids Res* **28**: 56-59.

Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Collado-Vides, J., Paley, S.M., Pellegrini-Toole, A., Bonavides, C., and Gama-Castro, S. (2002) The EcoCyc Database. *Nucleic Acids Res* **30**: 56-58.

Kerner, M.J., Naylor, D.J., Ishihama, Y., Maier, T., Chang, H.C., Stines, A.P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., and Hartl, F.U. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**: 209-220.

Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M., and Karp, P.D. (2005) EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res* **33**: D334-337.

Kim, Y.H., Han, K.Y., Lee, K., and Lee, J. (2005) Proteome response of *Escherichia coli* fed-batch culture to temperature downshift. *Appl Microbiol Biotechnol* **68**: 786-793.

Kramer, M., Kecskes, E., and Horvath, I. (1981) Guanosine polyphosphate production of *Escherichia coli* stringent and relaxed strains in the stationary phase of growth. *Acta Microbiol Acad Sci Hung* **28**: 165-170.

Krause, E., Wenschuh, H., and Jungblut, P.R. (1999) The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal Chem* **71**: 4160-4165.

Lacour, S., and Landini, P. (2004) SigmaS-dependent gene expression at the onset of stationary phase in *Escherichia coli*: function of sigmaS-dependent genes and

identification of their promoter sequences. *J Bacteriol* **186**: 7186-7195.

Lange, R., Fischer, D., and Hengge-Aronis, R. (1995) Identification of transcriptional start sites and the role of ppGpp in the expression of rpoS, the structural gene for the sigma S subunit of RNA polymerase in *Escherichia coli*. *J Bacteriol* **177**: 4676-4680.

Langen, H., Takacs, B., Evers, S., Berndt, P., Lahm, H.W., Wipf, B., Gray, C., and Fountoulakis, M. (2000) Two-dimensional map of the proteome of Haemophilus influenzae. *Electrophoresis* **21**: 411-429.

Larsen, M.R., and Roepstorff, P. (2000) Mass spectrometric identification of proteins and characterization of their post-translational modifications in proteome analysis. *Fresenius J Anal Chem* **366**: 677-690.

Le Bihan, T., Pinto, D., and Figeys, D. (2001) Nanoflow gradient generator coupled with mu-LC-ESI-MS/MS for protein identification. *Anal Chem* **73**: 1307-1315.

Le Bihan, T., Duewel, H.S., and Figeys, D. (2003) On-line strong cation exchange micro-HPLC-ESI-MS/MS for protein identification and process optimization. *J Am Soc Mass Spectrom* **14**: 719-727.

Leyer, G.J., Wang, L.L., and Johnson, E.A. (1995) Acid adaptation of *Escherichia coli* O157:H7 increases survival in acidic foods. *Appl Environ Microbiol* **61**: 3752-3755.

Lim, C.J., Daws, T., Gerami-Nejad, M., and Fuchs, J.A. (2000) Growth-phase regulation of the *Escherichia coli* thioredoxin gene. *Biochim Biophys Acta* **1491**: 1-6.

Link, A.J., Robison, K., and Church, G.M. (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* **18**: 1259-1313.

Loewen, P.C., and Hengge-Aronis, R. (1994) The role of the sigma factor sigma S (KatF) in bacterial global regulation. *Annu Rev Microbiol* **48**: 53-80.

Loo, R.R., Cavalcoli, J.D., VanBogelen, R.A., Mitchell, C., Loo, J.A., Moldover, B., and Andrews, P.C. (2001) Virtual 2-D gel electrophoresis: visualization and analysis of the E. coli proteome by mass spectrometry. *Anal Chem* **73**: 4063-4070.

Lopez-Campistrous, A., Semchuk, P., Burke, L., Palmer-Stone, T., Brokx, S.J., Broderick, G., Bottorff, D., Bolch, S., Weiner, J.H., and Ellison, M.J. (2005) Localization, annotation, and comparison of the *Escherichia coli* K-12 proteome under two states of growth. *Mol Cell Proteomics* **4**: 1205-1209.

Lottspeich, F. (1999) Proteome Analysis: A Pathway to the Functional Analysis of Proteins. *Angew Chem Int Ed Engl* **38**: 2476-2492.

Lukashin, A.V., and Borodovsky, M. (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26**: 1107-1115.

Lupas, A. (1997) Predicting coiled-coil regions in proteins. *Curr Opin Struct Biol* **7**: 388-393.

Marino-Ramirez, L., and Hu, J.C. (2002) Isolation and mapping of self-assembling protein domains encoded by the Saccharomyces cerevisiae genome using lambda repressor fusions. *Yeast* **19**: 641-650.

Marino-Ramirez, L., Campbell, L., and Hu, J.C. (2003) Screening peptide/protein libraries fused to the lambda repressor DNA-binding domain in E. coli cells. *Methods Mol Biol* **205**: 235-250.

Markov, D., Christie, G.E., Sauer, B., Calendar, R., Park, T., Young, R., and Severinov, K. (2004) P2 growth restriction on an rpoC mutant is suppressed by alleles of the Rz1

homolog lysC. *J Bacteriol* **186**: 4628-4637.

Martin, R.G., and Rosner, J.L. (2002) Genomics of the marA/soxS/rob regulon of *Escherichia coli*: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. *Mol Microbiol* **44**: 1611-1624.

Medigue, C., Viari, A., Henaut, A., and Danchin, A. (1993) Colibri: a functional data base for the *Escherichia coli* genome. *Microbiol Rev* **57**: 623-654.

Metcalf, W.W., Steed, P.M., and Wanner, B.L. (1990) Identification of phosphate starvation-inducible genes in *Escherichia coli* K-12 by DNA sequence analysis of psi::lacZ(Mu d1) transcriptional fusions. *J Bacteriol* **172**: 3191-3200.

Misra, R.V., Horler, R.S., Reindl, W., Goryanin, II, and Thomas, G.H. (2005) EchoBASE: an integrated post-genomic database for *Escherichia coli*. *Nucleic Acids Res* **33**: D329-333.

Mori, H. (2004) From the sequence to cell modeling: comprehensive functional genomics in *Escherichia coli*. *J Biochem Mol Biol* **37**: 83-92.

Mujacic, M., and Baneyx, F. (2007) Chaperone Hsp31 Contributes to Acid Resistance in Stationary-Phase *Escherichia coli*. *Appl Environ Microbiol* **73**: 1014-1018.

Muller, A., MacCallum, R.M., and Sternberg, M.J. (1999) Benchmarking PSI-BLAST in genome annotation. *J Mol Biol* **293**: 1257-1271.

Nash, R., Weng, S., Hitz, B., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Hong, E.L., Livstone, M.S., Oughtred, R., Park, J., Skrzypek, M., Theesfeld, C.L., Binkley, G., Dong, Q., Lane, C., Miyasato, S., Sethuraman, A., Schroeder, M., Dolinski, K., Botstein, D., and Cherry, J.M. (2007)

Expanded protein information at SGD: new pages and proteome browser. *Nucleic Acids Res* **35**: D468-471.

Nouwens, A.S., Cordwell, S.J., Larsen, M.R., Molloy, M.P., Gillings, M., Willcox, M.D., and Walsh, B.J. (2000) Complementing genomics with proteomics: the membrane subproteome of Pseudomonas aeruginosa PAO1. *Electrophoresis* **21**: 3797-3809.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**: 4007-4021.

Pachkov, M., Erb, I., Molina, N., and van Nimwegen, E. (2007) SwissRegulon: a database of genome-wide annotations of regulatory sites. *Nucleic Acids Res* **35**: D127-131.

Paley, S.M., and Karp, P.D. (1996) Adapting EcoCyc for use on the World Wide Web. *Gene* **172**: GC43-50.

Pappin, D.J., Hojrup, P., and Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3**: 327-332.

Partridge, J.D., Scott, C., Tang, Y., Poole, R.K., and Green, J. (2006) *Escherichia coli* transcriptome dynamics during the transition from anaerobic to aerobic conditions. *J Biol Chem* **281**: 27806-27815.

Patten, C.L., Kirchhof, M.G., Schertzberg, M.R., Morton, R.A., and Schellhorn, H.E. (2004) Microarray analysis of RpoS-mediated gene expression in *Escherichia coli* K-12. *Mol Genet Genomics* **272**: 580-591.

Pertsemlidis, A., and Fondon, J.W., 3rd (2001) Having a BLAST with bioinformatics (and avoiding BLASTphemy). *Genome Biol* **2**: REVIEWS2002.

Potamitou, A., Neubauer, P., Holmgren, A., and Vlamis-Gardikas, A. (2002) Expression of

Escherichia coli glutaredoxin 2 is mainly regulated by ppGpp and sigmaS. *J Biol Chem* **277**: 17775-17780.

Ramagopal, S., and Davis, B.D. (1974) Localization of the stringent protein of *Escherichia coli* on the 50S ribosomal subunit. *Proc Natl Acad Sci U S A* **71**: 820-824.

Raman, B., Nandakumar, M.P., Muthuvijayan, V., and Marten, M.R. (2005) Proteome analysis to assess physiological changes in *Escherichia coli* grown under glucose-limited fed-batch conditions. *Biotechnol Bioeng* **92**: 384-392.

Ramirez Santos, J., Contreras Ferrat, G., and Gomez Eichelmann, M.C. (2005) [Stationary phase in *Escherichia coli*]. *Rev Latinoam Microbiol* **47**: 92-101.

Remm, M., and Sonnhammer, E. (2000) Classification of transmembrane protein families in the Caenorhabditis elegans genome and identification of human orthologs. *Genome Res* **10**: 1679-1689.

Rey, S., Gardy, J.L., and Brinkman, F.S. (2005) Assessing the precision of high-throughput computational and laboratory approaches for the genome-wide identification of protein subcellular localization in bacteria. *BMC Genomics* **6**: 162.

Riehle, M.M., Bennett, A.F., and Long, A.D. (2005) Changes in gene expression following high-temperature adaptation in experimentally evolved populations of E. coli. *Physiol Biochem Zool* **78**: 299-315.

Riley, M., and Space, D.B. (1996) Genes and proteins of *Escherichia coli* (GenProtEc). *Nucleic Acids Res* **24**: 40.

Riley, M. (1997) Genes and proteins of *Escherichia coli* K-12 (GenProtEC). *Nucleic Acids Res* **25**: 51-52.

Riley, M., and Serres, M.H. (2000) Interim report on genomics of *Escherichia coli*. *Annu Rev Microbiol* **54**: 341-411.

Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., Mori, H., Perna, N.T., Plunkett, G., 3rd, Rudd, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D., and Wanner, B.L. (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Res* **34**: 1-9.

Roberts, M.D., Martin, N.L., and Kropinski, A.M. (2004) The genome and proteome of coliphage T1. *Virology* **318**: 245-266.

Robison, K., McGuire, A.M., and Church, G.M. (1998) A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J Mol Biol* **284**: 241-254.

Romeo, T., and Preiss, J. (1989) Genetic regulation of glycogen biosynthesis in *Escherichia coli*: in vitro effects of cyclic AMP and guanosine 5'-diphosphate 3'-diphosphate and analysis of in vivo transcripts. *J Bacteriol* **171**: 2773-2782.

Roseman, S., and Meadow, N.D. (1990) Signal transduction by the bacterial phosphotransferase system. Diauxie and the crr gene (J. Monod revisited). *J Biol Chem* **265**: 2993-2996.

Rosen, R., and Ron, E.Z. (2002) Proteome analysis in the study of the bacterial heat-shock response. *Mass Spectrom Rev* **21**: 244-265.

Rudd, K.E. (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res* **28**: 60-64.

Salgado, H., Santos, A., Garza-Ramos, U., van Helden, J., Diaz, E., and Collado-Vides, J. (1999) RegulonDB (version 2.0): a database on transcriptional regulation in *Escherichia coli*. *Nucleic Acids Res* **27**: 59-60.

Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Millan-Zarate, D., Blattner, F.R., and Collado-Vides, J. (2000) RegulonDB (version 3.0): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res* **28**: 65-67.

Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Millan-Zarate, D., Diaz-Peredo, E., Sanchez-Solano, F., Perez-Rueda, E., Bonavides-Martinez, C., and Collado-Vides, J. (2001) RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res* **29**: 72-74.

Salgado, H., Gama-Castro, S., Martinez-Antonio, A., Diaz-Peredo, E., Sanchez-Solano, F., Peralta-Gil, M., Garcia-Alonso, D., Jimenez-Jacinto, V., Santos-Zavaleta, A., Bonavides-Martinez, C., and Collado-Vides, J. (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res* **32**: D303-306.

Salgado, H., Gama-Castro, S., Peralta-Gil, M., Diaz-Peredo, E., Sanchez-Solano, F., Santos-Zavaleta, A., Martinez-Flores, I., Jimenez-Jacinto, V., Bonavides-Martinez, C., Segura-Salazar, J., Martinez-Antonio, A., and Collado-Vides, J. (2006) RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res* **34**: D394-397.

Schagger, H., and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal*

Biochem **166**: 368-379.

Schaumburg, J., Diekmann, O., Hagendorff, P., Bergmann, S., Rohde, M., Hammerschmidt, S., Jansch, L., Wehland, J., and Karst, U. (2004) The cell wall subproteome of Listeria monocytogenes. *Proteomics* **4**: 2991-3006.

Schembri, M.A., Kjaergaard, K., and Klemm, P. (2003) Global gene expression in *Escherichia coli* biofilms. *Mol Microbiol* **48**: 253-267.

Schmidt, C., Velleman, M., and Arber, W. (1996) Three functions of bacteriophage P1 involved in cell lysis. *J Bacteriol* **178**: 1099-1104.

Schreiber, G., Ron, E.Z., and Glaser, G. (1995) ppGpp-mediated regulation of DNA replication and cell division in *Escherichia coli*. *Curr Microbiol* **30**: 27-32.

Sendak, R.A., and Bensadoun, A. (1998) Identification of a heparin-binding domain in the distal carboxyl-terminal region of lipoprotein lipase by site-directed mutagenesis. *J Lipid Res* **39**: 1310-1315.

Serres, M.H., Goswami, S., and Riley, M. (2004) GenProtEC: an updated and improved analysis of functions of *Escherichia coli* K-12 proteins. *Nucleic Acids Res* **32**: D300-302.

Sharp, P.M., and Li, W.H. (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**: 1281-1295.

Shultz, J.L., Kurunam, D., Shopinski, K., Iqbal, M.J., Kazi, S., Zobrist, K., Bashir, R., Yaegashi, S., Lavu, N., Afzal, A.J., Yesudas, C.R., Kassem, M.A., Wu, C., Zhang, H.B., Town, C.D., Meksem, K., and Lightfoot, D.A. (2006) The Soybean Genome

Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of Glycine max. *Nucleic Acids Res* **34**: D758-765.

Silva, J.C., Denny, R., Dorschel, C., Gorenstein, M.V., Li, G.Z., Richardson, K., Wall, D., and Geromanos, S.J. (2006) Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics* **5**: 589-607.

Simpson, J.C., and Pepperkok, R. (2003) Localizing the proteome. *Genome Biol* **4**: 240.

Song, J., Burrage, K., Yuan, Z., and Huber, T. (2006) Prediction of cis/trans isomerization in proteins using PSI-BLAST profiles and secondary structure information. *BMC Bioinformatics* **7**: 124.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., and Lewis, S. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res* **12**: 1599-1610.

Sundararaj, S., Guo, A., Habibi-Nazhad, B., Rouani, M., Stothard, P., Ellison, M., and Wishart, D.S. (2004) The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of *Escherichia coli*. *Nucleic Acids Res* **32**: D293-295.

Tamarit, J., Cabiscol, E., and Ros, J. (1998) Identification of the major oxidatively damaged proteins in *Escherichia coli* cells exposed to oxidative stress. *J Biol Chem* **273**: 3027-3032.

Tani, T.H., Khodursky, A., Blumenthal, R.M., Brown, P.O., and Matthews, R.G. (2002) Adaptation to famine: a family of stationary-phase genes revealed by microarray

analysis. *Proc Natl Acad Sci U S A* **99**: 13471-13476.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and Conway, T. (1999) Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* **181**: 6425-6440.

Thiede, B., Hohenwarter, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F., and Jungblut, P.R. (2005) Peptide mass fingerprinting. *Methods* **35**: 237-247.

Tilly, K., McKittrick, N., Zylicz, M., and Georgopoulos, C. (1983) The dnaK protein modulates the heat-shock response of *Escherichia coli*. *Cell* **34**: 641-646.

Tonella, L., Hoogland, C., Binz, P.A., Appel, R.D., Hochstrasser, D.F., and Sanchez, J.C. (2001) New perspectives in the *Escherichia coli* proteome investigation. *Proteomics* **1**: 409-423.

Trost, M., Wehmhoner, D., Karst, U., Dieterich, G., Wehland, J., and Jansch, L. (2005) Comparative proteome analysis of secretory proteins from pathogenic and nonpathogenic Listeria species. *Proteomics* **5**: 1544-1557.

Ueberle, B., Frank, R., and Herrmann, R. (2002) The proteome of the bacterium Mycoplasma pneumoniae: comparing predicted open reading frames to identified gene products. *Proteomics* **2**: 754-764.

Usuda, Y., and Kurahashi, O. (2005) Effects of deregulation of methionine biosynthesis on methionine excretion in *Escherichia coli*. *Appl Environ Microbiol* **71**: 3228-3234.

Verhoeckx, K.C., Bijlsma, S., Jespersen, S., Ramaker, R., Verheij, E.R., Witkamp, R.F., van der Greef, J., and Rodenburg, R.J. (2004) Characterization of anti-inflammatory compounds using transcriptomics, proteomics, and metabolomics in combination

with multivariate data analysis. *Int Immunopharmacol* **4**: 1499-1514.

Vollmer, M., Nagele, E., and Horth, P. (2003) Differential proteome analysis: two-dimensional nano-LC/MS of E. coli proteome grown on different carbon sources. *J Biomol Tech* **14**: 128-135.

Washburn, M.P., Wolters, D., and Yates, J.R., 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242-247.

Weber, A., Kogl, S.A., and Jung, K. (2006) Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in *Escherichia coli*. *J Bacteriol* **188**: 7165-7175.

Weichart, D., Querfurth, N., Dreger, M., and Hengge-Aronis, R. (2003) Global role for ClpP-containing proteases in stationary-phase adaptation of *Escherichia coli*. *J Bacteriol* **185**: 115-125.

Wendrich, T.M., Blaha, G., Wilson, D.N., Marahiel, M.A., and Nierhaus, K.H. (2002) Dissection of the mechanism for the stringent factor RelA. *Mol Cell* **10**: 779-788.

Wenschuh, H., Halada, P., Lamer, S., Jungblut, P., and Krause, E. (1998) The ease of peptide detection by matrix-assisted laser desorption/ionization mass spectrometry: the effect of secondary structure on signal intensity. *Rapid Commun Mass Spectrom* **12**: 115-119.

Westermeier, R., Postel, W., Weser, J., and Gorg, A. (1983) High-resolution two-dimensional electrophoresis with isoelectric focusing in immobilized pH gradients. *J Biochem Biophys Methods* **8**: 321-330.

Wiback, S.J., Mahadevan, R., and Palsson, B.O. (2004) Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the *Escherichia coli* spectrum. *Biotechnol Bioeng* **86**: 317-331.

Winsor, G.L., Lo, R., Sui, S.J., Ung, K.S., Huang, S., Cheng, D., Ching, W.K., Hancock, R.E., and Brinkman, F.S. (2005) Pseudomonas aeruginosa Genome Database and PseudoCAP: facilitating community-based, continually updated, genome annotation. *Nucleic Acids Res* **33**: D338-343.

Wolf, B., Lausarot, P.M., Lesnaw, J.A., and Reichmann, M.E. (1970) Preparation of polymeric protein markers and an investigation of their behavior in sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Biochim Biophys Acta* **200**: 180-183.

Wolters, D.A., Washburn, M.P., and Yates, J.R., 3rd (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**: 5683-5690.

Wu, C.C., and Yates, J.R., 3rd (2003) The application of mass spectrometry to membrane proteomics. *Nat Biotechnol* **21**: 262-267.

Xu, C., Lin, X., Ren, H., Zhang, Y., Wang, S., and Peng, X. (2006) Analysis of outer membrane proteome of *Escherichia coli* related to resistance to ampicillin and tetracycline. *Proteomics* **6**: 462-473.

Xu, M., Struck, D.K., Deaton, J., Wang, I.N., and Young, R. (2004) A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. *Proc Natl Acad Sci U S A* **101**: 6415-6420.

Xu, M., Arulandu, A., Struck, D.K., Swanson, S., Sacchettini, J.C., and Young, R. (2005) Disulfide isomerization after membrane release of its SAR domain activates P1

lysozyme. *Science* **307**: 113-117.

Yamanaka, K., and Inouye, M. (1997) Growth-phase-dependent expression of cspD, encoding a member of the CspA family in *Escherichia coli*. *J Bacteriol* **179**: 5126-5130.

Yan, J.X., Devenish, A.T., Wait, R., Stone, T., Lewis, S., and Fowler, S. (2002) Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of *Escherichia coli*. *Proteomics* **2**: 1682-1698.

Yang, X., and Ishiguro, E.E. (2001) Involvement of the N terminus of ribosomal protein L11 in regulation of the RelA protein of *Escherichia coli*. *J Bacteriol* **183**: 6532-6537.

Yates, J.R., 3rd, Speicher, S., Griffin, P.R., and Hunkapiller, T. (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem* **214**: 397-408.

Yates, J.R., 3rd (2000) Mass spectrometry. From genomics to proteomics. *Trends Genet* **16**: 5-8.

Yoon, S.H., Han, M.J., Lee, S.Y., Jeong, K.J., and Yoo, J.S. (2003) Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. *Biotechnol Bioeng* **81**: 753-767.

Zhang, H., Yan, W., and Aebersold, R. (2004) Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr Opin Chem Biol* **8**: 66-75.

Zhang, N., and Young, R. (1999) Complementation and characterization of the nested Rz and Rz1 reading frames in the genome of bacteriophage lambda. *Mol Gen Genet* **262**: 659-667.

Zhang, W., Culley, D.E., Nie, L., and Brockman, F.J. (2006) DNA microarray analysis of anaerobic Methanosarcina barkeri reveals responses to heat shock and air exposure. *J Ind Microbiol Biotechnol* **33**: 784-790.

Zhen, Y., Xu, N., Richardson, B., Becklin, R., Savage, J.R., Blake, K., and Peltier, J.M. (2004) Development of an LC-MALDI method for the analysis of protein complexes. *J Am Soc Mass Spectrom* **15**: 803-822.

APPENDIX A

Table A.1. Heparin-binding proteins identified in every fraction from heparin column.

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	1	*accC*	P24182	49.32	6.65	54	4
Exponential	1	*aceE*	P06958	99.54	5.46	110	
Exponential	1	*adhE*	P17547	96	6.33	110	
Exponential	1	*crl*	P24251	15.52	5.67	18	2
Exponential	1	*dnaK*	P04475	68.98	4.83	80	
Exponential	1	*fabB*	P14926	42.61	5.35	54;51	3
Exponential	1	*fabF*	P39435	42.91	5.71	51	
Exponential	1	*fabI*	P29132	27.73	5.58	37	6
Exponential	1	*ftnA*	P23887	19.42	4.77	18	
Exponential	1	*fur*	P06975	16.79	5.68	18	
Exponential	1	*glnS*	P00962	63.35	5.89	69	
Exponential	1	*gltX*	P04805	53.82	5.59	60	4
Exponential	1	*gnd*	P00350	51.48	5.05	48	
Exponential	1	*hisG*	P10366	33.37	5.47		3
Exponential	1	*lpd*	P00391	50.56	5.79	60	4
Exponential	1	*metE*	P25665	84.54	5.61	96;80;69	
Exponential	1	*metF*	P00394	33.1	6	48	
Exponential	1	*mprA*	P24201	20.56	5.82	21	2
Exponential	1	*ompC*	P06996	38.31	4.48	40	
Exponential	1	*ompT*	P09169	33.48	5.38	40	
Exponential	1	*oppA*	P23843	58.36	5.85	60	
Exponential	1	*pgk*	P11665	40.99	5.08		5
Exponential	1	*ppc*	P00864	99.06	5.52	96	
Exponential	1	*ppk*	P28688	80.3	8.96	43;40	
Exponential	1	*purU*	P37051	31.93	6.5	48	
Exponential	1	*rlpB*	P10101	19.46	8.05	60	
Exponential	1	*rplL*	P02392	12.16	4.6		1
Exponential	1	*rplP*	P02414	15.28	11.22		1
Exponential	1	*rplY*	P02426	10.69	9.6		3
Exponential	1	*rpsA*	P02349	61.16	4.89	80	
Exponential	1	*rpsT*	P02378	9.55	11.18		1
Exponential	1	*sucA*	P07015	105.06	6.04	110	
Exponential	1	*tufA*	P02990	43.18	5.3		4
Exponential	1	*yfjH*	P52123	37.14	7.04	40	
Exponential	1	*yggS*	P52054	25.79	6.09	30	
Exponential	1	*ykgM*	P71302	9.92	9.3		1

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	2	*accC*	P24182	49.32	6.65		3
Exponential	2	*adhE*	P17547	96	6.33	116	
Exponential	2	*crl*	P24251	15.52	5.67	18	
Exponential	2	*fabB*	P14926	42.61	5.35	150;50;40	5
Exponential	2	*fabG*	P25716	25.56	6.76	29	
Exponential	2	*fabH*	P24249	33.52	5.08	36	2
Exponential	2	*fabI*	P29132	27.73	5.58	36	
Exponential	2	*fur*	P06975	16.79	5.68	18	
Exponential	2	*gapA*	P06977	35.4	6.58	40	
Exponential	2	*glnS*	P00962	63.35	5.89	69	
Exponential	2	*gltB*	P09831	161.93	6.19	190	
Exponential	2	*gltX*	P04805	53.82	5.59	57	
Exponential	2	*hisG*	P10366	33.37	5.47	36	2
Exponential	2	*lpd*	P00391	50.56	5.79	57	2
Exponential	2	*metE*	P25665	84.54	5.61	150;80;69	2
Exponential	2	*metG*	P00959	76.12	5.56	80	
Exponential	2	*mfd*	P30958	129.98	5.79	150	
Exponential	2	*mprA*	P24201	20.56	5.82		3
Exponential	2	*ompA*	P02934	35.17	5.6	36	
Exponential	2	*ompC*	P06996	38.31	4.48	40	
Exponential	2	*ompF*	P02931	37.08	4.64	40	
Exponential	2	*ompR*	P03025	27.35	6.04	29	
Exponential	2	*ompT*	P09169	33.48	5.38	40	
Exponential	2	*oppA*	P23843	58.36	5.85	190;96;62;57	7
Exponential	2	*pbpC*	P76577	85.07	9.52	96	
Exponential	2	*pdxH*	P28225	25.41	9.18	29	
Exponential	2	*pgk*	P11665	40.99	5.08	190;150;44;29	15
Exponential	2	*pnp*	P05055	77.1	5.11	93	
Exponential	2	*rplC*	P02386	22.24	9.9		4
Exponential	2	*rplJ*	P02408	17.58	9.04		2
Exponential	2	*rplL*	P02392	12.16	4.6		3
Exponential	2	*rplU*	P02422	11.56	9.85		2
Exponential	2	*rplY*	P02426	10.69	9.6		5
Exponential	2	*sucA*	P07015	105.06	6.04	116	
Exponential	2	*thrS*	P00955	74.01	5.8	80	
Exponential	2	*trpS*	P00954	37.44	6.27	40	
Exponential	2	*ung*	P12295	25.56	6.71	29	
Exponential	2	*yggS*	P52054	25.79	6.09	29	
Exponential	2	*ykgM*	P71302	9.92	9.3		2

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	3	*accC*	P24182	49.32	6.65		5
Exponential	3	*aceE*	P06958	99.54	5.46	140	
Exponential	3	*adhE*	P17547	96	6.33	140	
Exponential	3	*crl*	P24251	15.52	5.67		4
Exponential	3	*fabB*	P14926	42.61	5.35	50;40	9
Exponential	3	*fabF*	P39435	42.91	5.71	50	
Exponential	3	*fabH*	P24249	33.52	5.08		5
Exponential	3	*fabI*	P29132	27.73	5.58		5
Exponential	3	*glgB*	P07762	84.34	5.91	90	
Exponential	3	*glnS*	P00962	63.35	5.89	80;70	5
Exponential	3	*gltB*	P09831	161.93	6.19	190	
Exponential	3	*gltX*	P04805	53.82	5.59	55	10
Exponential	3	*gnd*	P00350	51.48	5.05	50	
Exponential	3	*hisG*	P10366	33.37	5.47		4
Exponential	3	*lpd*	P00391	50.56	5.79	55	6
Exponential	3	*metE*	P25665	84.54	5.61	110;96;90;80	1
Exponential	3	*metG*	P00959	76.12	5.56	80	
Exponential	3	*minE*	P18198	10.23	5.15		3
Exponential	3	*mprA*	P24201	20.56	5.82	22	3
Exponential	3	*nusB*	P04381	15.69	6.6		2
Exponential	3	*ompR*	P03025	27.35	6.04		3
Exponential	3	*oppA*	P23843	58.36	5.85	190;80;70;60;55;50	8
Exponential	3	*pgk*	P11665	40.99	5.08	190;90;44;40;36;29;22	22
Exponential	3	*pnp*	P05055	77.1	5.11	90	
Exponential	3	*rplC*	P02386	22.24	9.9		4
Exponential	3	*rplD*	P02388	22.09	9.72		3
Exponential	3	*rplJ*	P02408	17.58	9.04		4
Exponential	3	*rplL*	P02392	12.16	4.6		5
Exponential	3	*rplS*	P02420	13	10.62		2
Exponential	3	*rplU*	P02422	11.56	9.85		2
Exponential	3	*rplY*	P02426	10.69	9.6		4
Exponential	3	*rpoA*	P00574	36.51	4.98		7
Exponential	3	*rpsH*	P02361	14	9.44		5
Exponential	3	*rpsJ*	P02364	11.74	9.68		1
Exponential	3	*sucA*	P07015	105.06	6.04	140	4
Exponential	3	*trpS*	P00954	37.44	6.27		1
Exponential	3	*tufA*	P02990	43.18	5.3	50	6
Exponential	3	*ykgM*	P71302	9.92	9.3		4

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	4	*accC*	P24182	49.32	6.65		5
Exponential	4	*aceE*	P06958	99.54	5.46		4
Exponential	4	*adhE*	P17547	96	6.33	110	16
Exponential	4	*crl*	P24251	15.52	5.67		2
Exponential	4	*era*	P06616	33.81	6.73	43	
Exponential	4	*fabB*	P14926	42.61	5.35	43;37	9
Exponential	4	*fabG*	P25716	25.56	6.76		3
Exponential	4	*fabH*	P24249	33.52	5.08		5
Exponential	4	*fabI*	P29132	27.73	5.58	33	
Exponential	4	*fur*	P06975	16.79	5.68		2
Exponential	4	*glgB*	P07762	84.34	5.91		2
Exponential	4	*glnS*	P00962	63.35	5.89	75;70	7
Exponential	4	*gltB*	P09831	161.93	6.19		11
Exponential	4	*gltX*	P04805	53.82	5.59		6
Exponential	4	*gnd*	P00350	51.48	5.05		12
Exponential	4	*hisG*	P10366	33.37	5.47		6
Exponential	4	*iscS*	P39171	45.09	5.94		7
Exponential	4	*lpd*	P00391	50.56	5.79		13
Exponential	4	*metE*	P25665	84.54	5.61	180;140	5
Exponential	4	*mfd*	P30958	129.98	5.79		4
Exponential	4	*minE*	P18198	10.23	5.15		2
Exponential	4	*mprA*	P24201	20.56	5.82	30	2
Exponential	4	*ompC*	P06996	38.31	4.48	40	
Exponential	4	*ompR*	P03025	27.35	6.04	43;40;37;33	5
Exponential	4	*oppA*	P23843	58.36	5.85	43;37	9
Exponential	4	*panC*	P31663	31.6	5.92	43	
Exponential	4	*pgk*	P11665	40.99	5.08	43;40;37;33;30	24
Exponential	4	*rbfA*	P09170	15.02	5.96		4
Exponential	4	*rplA*	P02384	24.6	9.64		2
Exponential	4	*rplC*	P02386	22.24	9.9		3
Exponential	4	*rplJ*	P02408	17.58	9.04		3
Exponential	4	*rplL*	P02392	12.16	4.6		4
Exponential	4	*rplS*	P02420	13	10.62		2
Exponential	4	*rplU*	P02422	11.56	9.85		2
Exponential	4	*rplY*	P02426	10.69	9.6		6
Exponential	4	*rpoA*	P00574	36.51	4.98	60	6
Exponential	4	*rpsH*	P02361	14	9.44		3
Exponential	4	*rpsJ*	P02364	11.74	9.68		2

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	4	*rpsS*	P02375	10.3	10.52		1
Exponential	4	*slyA*	P55740	16.35	6.6		2
Exponential	4	*sucA*	P07015	105.06	6.04	110	23
Exponential	4	*sucB*	P07016	43.88	5.58	52	11
Exponential	4	*thrS*	P00955	74.01	5.8		7
Exponential	4	*trpS*	P00954	37.44	6.27	40	6
Exponential	4	*tufA*	P02990	43.18	5.3		6
Exponential	4	*ung*	P12295	25.56	6.71	37	
Exponential	4	*yciK*	P31808	27.93	7.67		2
Exponential	4	*yggS*	P52054	25.79	6.09	37	
Exponential	4	*yhaJ*	P42623	33.26	6.05		4
Exponential	4	*yibQ*	P37691	32.87	9.45	30	
Exponential	4	*yjeQ*	P39286	36.83	5.09		3
Exponential	4	*ykgM*	P71302	9.92	9.3		2
Exponential	5	*aceE*	P06958	99.54	5.46		4
Exponential	5	*adhE*	P17547	96	6.33	115	9
Exponential	5	*def*	P27251	19.2	5.23		2
Exponential	5	*dksA*	P18274	17.53	5.06		2
Exponential	5	*fabB*	P14926	42.61	5.35		3
Exponential	5	*fabG*	P25716	25.56	6.76	32;25	2
Exponential	5	*fabH*	P24249	33.52	5.08		3
Exponential	5	*fur*	P06975	16.79	5.68		1
Exponential	5	*ghrA*	P75913	36.84	6.76	29	3
Exponential	5	*gidB*	P17113	23.43	6.06	32	
Exponential	5	*glnS*	P00962	63.35	5.89	66	7
Exponential	5	*gltB*	P09831	161.93	6.19	180	22
Exponential	5	*gnd*	P00350	51.48	5.05	55;46	19
Exponential	5	*guaB*	P06981	52.02	6.02		11
Exponential	5	*hisG*	P10366	33.37	5.47		5
Exponential	5	*hscC*	P77319	61.99	5.05		2
Exponential	5	*iscS*	P39171	45.09	5.94		7
Exponential	5	*lpd*	P00391	50.56	5.79		5
Exponential	5	*lpxB*	P10441	42.38	6.57	38	
Exponential	5	*metE*	P25665	84.54	5.61	96	5
Exponential	5	*metF*	P00394	33.1	6	35	1
Exponential	5	*mprA*	P24201	20.56	5.82		2
Exponential	5	*ompA*	P02934	35.17	5.6	29	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	5	*oppA*	P23843	58.36	5.85	180;70;66;55;46;36; 35;32;29;28	5
Exponential	5	*parE*	P20083	70.24	5.44	26	
Exponential	5	*pgk*	P11665	40.99	5.08	43;36;35;32	12
Exponential	5	*rbfA*	P09170	15.02	5.96	17	5
Exponential	5	*rplJ*	P02408	17.58	9.04	17	3
Exponential	5	*rplL*	P02392	12.16	4.6	17	4
Exponential	5	*rplS*	P02420	13	10.62	17	4
Exponential	5	*rplY*	P02426	10.69	9.6		3
Exponential	5	*rpoA*	P00574	36.51	4.98		5
Exponential	5	*rpsH*	P02361	14	9.44		1
Exponential	5	*skp*	P11457	15.69	9.52	21	
Exponential	5	*sucA*	P07015	105.06	6.04	115	9
Exponential	5	*sucB*	P07016	43.88	5.58	55	7
Exponential	5	*thrS*	P00955	74.01	5.8	70	13
Exponential	5	*trpS*	P00954	37.44	6.27	38	8
Exponential	5	*tufA*	P02990	43.18	5.3		6
Exponential	5	*ycgK*	P76002	12.52	9.01	17	2
Exponential	5	*yegQ*	P76403	51.19	5.8		4
Exponential	5	*yggS*	P52054	25.79	6.09	32	
Exponential	5	*ykgM*	P71302	9.92	9.3		2
Exponential	5	*ynhG*	P76193	33.67	9.21		2
Exponential	5	*zwf*	P22992	55.7	5.56		3
Exponential	6	*accA*	P30867	35.11	5.76	37;35	17
Exponential	6	*accD*	P08193	33.32	7.58		8
Exponential	6	*aceE*	P06958	99.54	5.46		22
Exponential	6	*adhE*	P17547	96	6.33	68	9
Exponential	6	*alaS*	P00957	96.03	5.53		8
Exponential	6	*argA*	P08205	49.2	6.09		5
Exponential	6	*aroH*	P00887	38.74	6.42		9
Exponential	6	*atpA*	P00822	55.22	5.8		7
Exponential	6	*atpD*	P00824	50.19	4.9		11
Exponential	6	*bglX*	P33363	81.41	5.77	52	
Exponential	6	*dapE*	P24176	41.27	5.31	52	
Exponential	6	*def*	P27251	19.2	5.23		4
Exponential	6	*dinG*	P27296	81.44	7.62		5
Exponential	6	*dksA*	P18274	17.53	5.06		4

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	6	*erfK*	P39176	32.08	7.97		7
Exponential	6	*fabA*	P18391	18.84	6.16		2
Exponential	6	*fabG*	P25716	25.56	6.76		11
Exponential	6	*fadR*	P09371	26.84	6.54		3
Exponential	6	*fur*	P06975	16.79	5.68		3
Exponential	6	*ghrA*	P75913	36.84	6.76	37;35	15
Exponential	6	*glgB*	P07762	84.34	5.91		3
Exponential	6	*glmU*	P17114	49.19	6.09		7
Exponential	6	*glnE*	P30870	108.42	5.08		6
Exponential	6	*glnS*	P00962	63.35	5.89		23
Exponential	6	*gltB*	P09831	161.93	6.19		49
Exponential	6	*glyA*	P00477	45.32	6.03		6
Exponential	6	*gnd*	P00350	51.48	5.05	52;46;40;35;28;25	35
Exponential	6	*guaB*	P06981	52.02	6.02	60	21
Exponential	6	*gyrA*	P09097	96.96	5.09		7
Exponential	6	*gyrB*	P06982	89.82	5.72		9
Exponential	6	*hns*	P08936	15.41	5.44		4
Exponential	6	*infB*	P02995	97.35	5.8	68;52	41
Exponential	6	*iscS*	P39171	45.09	5.94	43	20
Exponential	6	*ligA*	P15042	73.61	5.39		5
Exponential	6	*lpd*	P00391	50.56	5.79		5
Exponential	6	*map*	P07906	29.33	5.64		3
Exponential	6	*metE*	P25665	84.54	5.61	68;60;57;52	43
Exponential	6	*metF*	P00394	33.1	6	35	15
Exponential	6	*mfd*	P30958	129.98	5.79		14
Exponential	6	*miaA*	P16384	35.07	5.68		7
Exponential	6	*mnmG*	P17112	69.52	6.2		14
Exponential	6	*nuoG*	P33602	100.17	5.85		7
Exponential	6	*ompA*	P02934	35.17	5.6		4
Exponential	6	*oppA*	P23843	58.36	5.85	60;57	8
Exponential	6	*pdxH*	P28225	25.41	9.18		4
Exponential	6	*pgk*	P11665	40.99	5.08	43;33	13
Exponential	6	*pheS*	P08312	36.83	5.79		3
Exponential	6	*pntA*	P07001	54.62	5.65		5
Exponential	6	*prc*	P23865	74.32	6.04		7
Exponential	6	*proC*	P00373	28.14	5.64		3
Exponential	6	*purU*	P37051	31.93	6.5	38	10

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	6	*rbfA*	P09170	15.02	5.96	16	7
Exponential	6	*relA*	P11585	83.88	6.29		15
Exponential	6	*rho*	P03002	47	6.75		9
Exponential	6	*rne*	P21513	118.2	5.48		7
Exponential	6	*rplA*	P02384	24.6	9.64		8
Exponential	6	*rplB*	P02387	29.73	10.93		5
Exponential	6	*rplC*	P02386	22.24	9.9		6
Exponential	6	*rplF*	P02390	18.77	9.71		6
Exponential	6	*rplI*	P02418	15.77	6.17		4
Exponential	6	*rplJ*	P02408	17.58	9.04		3
Exponential	6	*rplK*	P02409	14.74	9.64		2
Exponential	6	*rplL*	P02392	12.16	4.6		3
Exponential	6	*rplM*	P02410	16.02	9.91		3
Exponential	6	*rplN*	P02411	13.54	10.43		2
Exponential	6	*rplO*	P02413	14.98	11.18		6
Exponential	6	*rplS*	P02420	13	10.62		6
Exponential	6	*rplU*	P02422	11.56	9.85		3
Exponential	6	*rplY*	P02426	10.69	9.6		6
Exponential	6	*rpmA*	P02427	8.99	10.58		3
Exponential	6	*rpoA*	P00574	36.51	4.98		7
Exponential	6	*rpoB*	P00575	150.63	5.15		13
Exponential	6	*rpoD*	P00579	70.26	4.69		9
Exponential	6	*rpsH*	P02361	14	9.44		3
Exponential	6	*rpsJ*	P02364	11.74	9.68		6
Exponential	6	*rpsN*	P02370	11.45	11.16		2
Exponential	6	*rpsO*	P02371	10.14	10.4		2
Exponential	6	*rpsQ*	P02373	9.57	9.64		2
Exponential	6	*rpsU*	P02379	8.37	11.15		4
Exponential	6	*rsd*	P31690	18.24	5.65		4
Exponential	6	*skp*	P11457	15.69	9.52	20	8
Exponential	6	*sucA*	P07015	105.06	6.04	68	7
Exponential	6	*sucB*	P07016	43.88	5.58		9
Exponential	6	*talB*	P30148	35.09	5.11	37	
Exponential	6	*thrS*	P00955	74.01	5.8		29
Exponential	6	*tktA*	P27302	72.21	5.43		4
Exponential	6	*trpR*	P03032	12.22	5.42		2
Exponential	6	*trpS*	P00954	37.44	6.27		7
Exponential	6	*tufA*	P02990	43.18	5.3		13
Exponential	6	*tyrR*	P07604	57.66	5.54		11
Exponential	6	*tyrS*	P00951	47.4	5.59		13
Exponential	6	*ugpQ*	P10908	27.41	6.09		5
Exponential	6	*uidR*	Q59431	21.8	5.69	28	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	6	*uvrB*	P07025	76.09	5.14		4
Exponential	6	*uvrY*	P07027	23.89	6.53		4
Exponential	6	*xthA*	P09030	30.97	5.8		6
Exponential	6	*yafC*	P30864	33.78	6.93		5
Exponential	6	*yafJ*	Q47147	28.64	8.14		5
Exponential	6	*ybaK*	P37175	17.09	9.02		3
Exponential	6	*ybeZ*	P77349	40.65	6.24		14
Exponential	6	*ybgK*	P75745	34.39	8.59		3
Exponential	6	*ycgK*	P76002	12.52	9.01	33;16	10
Exponential	6	*ycjX*	P76046	52.61	8.25		3
Exponential	6	*ydgH*	P76177	31.91	9.1	37	14
Exponential	6	*yegQ*	P76403	51.19	5.8		11
Exponential	6	*yfaA*	P17994	62.67	5.32		3
Exponential	6	*ykgM*	P71302	9.92	9.3		8
Exponential	6	*ynhG*	P76193	33.67	9.21		11
Exponential	6	*zwf*	P22992	55.7	5.56		18
Exponential	7	*accA*	P30867	35.11	5.76	34	6
Exponential	7	*fabA*	P18391	18.84	6.16	20	
Exponential	7	*fabG*	P25716	25.56	6.76	29;27	
Exponential	7	*fadR*	P09371	26.84	6.54	30	
Exponential	7	*gnd*	P00350	51.48	5.05	50	
Exponential	7	*infB*	P02995	97.35	5.8		5
Exponential	7	*iscS*	P39171	45.09	5.94	50;42	8
Exponential	7	*map*	P07906	29.33	5.64		1
Exponential	7	*metE*	P25665	84.54	5.61	90;80;70;66;60;56;53;50;45;42;39;37;34;30;29;27;25	30
Exponential	7	*metJ*	P08338	12.01	5.4		2
Exponential	7	*purU*	P37051	31.93	6.5	34;30	
Exponential	7	*rcsB*	P14374	23.67	6.85	29	
Exponential	7	*rplK*	P02409	14.74	9.64		2
Exponential	7	*rplO*	P02413	14.98	11.18		4
Exponential	7	*rplS*	P02420	13	10.62		3
Exponential	7	*rpoA*	P00574	36.51	4.98		3
Exponential	7	*rpsN*	P02370	11.45	11.16		1
Exponential	7	*skp*	P11457	15.69	9.52	20	
Exponential	7	*xthA*	P09030	30.97	5.8	34	
Exponential	7	*ycgK*	P76002	12.52	9.01		3
Exponential	7	*ydgH*	P76177	31.91	9.1		6

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	8	*arcA*	P03026	27.29	5.21	32;29	5
Exponential	8	*aroC*	P12008	39.01	5.82	40	
Exponential	8	*fabG*	P25716	25.56	6.76	29	4
Exponential	8	*gyrA*	P09097	96.96	5.09	230	7
Exponential	8	*hisS*	P04804	46.9	5.65	46	6
Exponential	8	*hns*	P08936	15.41	5.44		2
Exponential	8	*infA*	P02998	8.12	9.23		3
Exponential	8	*iscS*	P39171	45.09	5.94	44	4
Exponential	8	*map*	P07906	29.33	5.64		2
Exponential	8	*mdoG*	P33136	55.37	6.26	48	
Exponential	8	*metE*	P25665	84.54	5.61	230;170;80;55;48 ;46;40;29;18	24
Exponential	8	*metJ*	P08338	12.01	5.4		4
Exponential	8	*mug*	P43342	18.67	9.17	21	
Exponential	8	*rho*	P03002	47	6.75	46	
Exponential	8	*rplA*	P02384	24.6	9.64	32;29	4
Exponential	8	*rplF*	P02390	18.77	9.71	21	7
Exponential	8	*rplI*	P02418	15.77	6.17	18	3
Exponential	8	*rplJ*	P02408	17.58	9.04	18	
Exponential	8	*rplO*	P02413	14.98	11.18		4
Exponential	8	*rplS*	P02420	13	10.62		2
Exponential	8	*rplX*	P02425	11.19	10.21		3
Exponential	8	*rpoA*	P00574	36.51	4.98	40	3
Exponential	8	*rpoB*	P00575	150.63	5.15	170	
Exponential	8	*rpsB*	P02351	26.61	6.69	32	
Exponential	8	*rpsJ*	P02364	11.74	9.68		3
Exponential	8	*rpsN*	P02370	11.45	11.16		2
Exponential	8	*rpsQ*	P02373	9.57	9.64		2
Exponential	8	*rpsS*	P02375	10.3	10.52		2
Exponential	8	*stpA*	P30017	15.35	7.95	18	
Exponential	8	*suhB*	P22783	29.17	6.45	32	5
Exponential	8	*tgt*	P19675	42.59	5.97	18	
Exponential	8	*ycbB*	P22525	67.81	8.63	55	
Exponential	8	*ydcC*	P28917	43.08	6.97	29	
Exponential	8	*yibK*	P33899	17.73	6.2	21	
Exponential	8	*yjgA*	P26650	21.36	5.3	21	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	9	*arcA*	P03026	27.29	5.21		3
Exponential	9	*aroC*	P12008	39.01	5.82	43	
Exponential	9	*atpD*	P00824	50.19	4.9	46	
Exponential	9	*creA*	P08367	17.11	9.04		3
Exponential	9	*crp*	P03020	23.64	8.38	29	4
Exponential	9	*glgC*	P00584	48.57	5.78	25	
Exponential	9	*gltB*	P09831	161.93	6.19	25	
Exponential	9	*gltX*	P04805	53.82	5.59	25	
Exponential	9	*gyrA*	P09097	96.96	5.09		7
Exponential	9	*hflX*	P25519	48.33	5.68	47	
Exponential	9	*hisS*	P04804	46.9	5.65	47;46	10
Exponential	9	*hns*	P08936	15.41	5.44		3
Exponential	9	*lexA*	P03033	22.36	6.23	29	
Exponential	9	*lrp*	P19494	18.76	8.91		5
Exponential	9	*mdoG*	P33136	55.37	6.26	66	
Exponential	9	*metE*	P25665	84.54	5.61	90;66	7
Exponential	9	*moaB*	P30746	18.53	5.73		5
Exponential	9	*oppA*	P23843	58.36	5.85	25;23	
Exponential	9	*oxyR*	P11721	34.28	5.96	43	
Exponential	9	*pepA*	P11648	54.88	6.82	47	3
Exponential	9	*pgk*	P11665	40.99	5.08	23	
Exponential	9	*rho*	P03002	47	6.75	47;46	12
Exponential	9	*rluE*	P75966	24.88	10.03	29	
Exponential	9	*rne*	P21513	118.2	5.48	66	
Exponential	9	*rplA*	P02384	24.6	9.64	31;30	6
Exponential	9	*rplB*	P02387	29.73	10.93		4
Exponential	9	*rplC*	P02386	22.24	9.9		7
Exponential	9	*rplE*	P02389	20.17	9.49	26	8
Exponential	9	*rplF*	P02390	18.77	9.71	26;25;21	11
Exponential	9	*rplI*	P02418	15.77	6.17	18	11
Exponential	9	*rplJ*	P02408	17.58	9.04	18	5
Exponential	9	*rplL*	P02392	12.16	4.6		3
Exponential	9	*rplM*	P02410	16.02	9.91		3
Exponential	9	*rplO*	P02413	14.98	11.18		4
Exponential	9	*rplS*	P02420	13	10.62		3
Exponential	9	*rplX*	P02425	11.19	10.21		5
Exponential	9	*rpoA*	P00574	36.51	4.98		4
Exponential	9	*rpoB*	P00575	150.63	5.15		6
Exponential	9	*rpsB*	P02351	26.61	6.69	31;30;29	9
Exponential	9	*rpsJ*	P02364	11.74	9.68		4
Exponential	9	*rpsQ*	P02373	9.57	9.64		3

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	9	*secA*	P10408	102.02	5.43	116;90;66	11
Exponential	9	*stpA*	P30017	15.35	7.95		4
Exponential	9	*suhB*	P22783	29.17	6.45		4
Exponential	9	*ttk*	P06969	22.84	8.78	31	
Exponential	9	*ydfH*	P77577	26.57	5.91	30	
Exponential	9	*ydjA*	P24250	20.06	6.31		3
Exponential	9	*yfiF*	P33635	37.78	8.94	43	
Exponential	9	*ykgM*	P71302	9.92	9.3		3
Exponential	10	*ccmH*	P33925	37.18	6.22	40	
Exponential	10	*crp*	P03020	23.64	8.38	26	6
Exponential	10	*cysH*	P17854	27.84	5.51	28	4
Exponential	10	*gyrA*	P09097	96.96	5.09		4
Exponential	10	*hflX*	P25519	48.33	5.68		4
Exponential	10	*lon*	P08177	87.44	6.01	17	
Exponential	10	*metE*	P25665	84.54	5.61		1
Exponential	10	*moaB*	P30746	18.53	5.73	36	2
Exponential	10	*moaC*	P30747	17.34	6.58	28	
Exponential	10	*oxyR*	P11721	34.28	5.96	40	
Exponential	10	*rplB*	P02387	29.73	10.93		4
Exponential	10	*rplC*	P02386	22.24	9.9		5
Exponential	10	*rplD*	P02388	22.09	9.72	17	1
Exponential	10	*rplE*	P02389	20.17	9.49	22	11
Exponential	10	*rplF*	P02390	18.77	9.71	36;23	7
Exponential	10	*rplI*	P02418	15.77	6.17	40;17	10
Exponential	10	*rplJ*	P02408	17.58	9.04	17	5
Exponential	10	*rplL*	P02392	12.16	4.6		4
Exponential	10	*rplM*	P02410	16.02	9.91		1
Exponential	10	*rplO*	P02413	14.98	11.18		5
Exponential	10	*rplS*	P02420	13	10.62		2
Exponential	10	*rplX*	P02425	11.19	10.21		3
Exponential	10	*rpmB*	P02428	8.88	11.42		2
Exponential	10	*rpmC*	P02429	7.27	9.98		2
Exponential	10	*rpoA*	P00574	36.51	4.98	47	7
Exponential	10	*rpoB*	P00575	150.63	5.15	28	14
Exponential	10	*rpoC*	P00577	155.16	6.67	28	11
Exponential	10	*rpsB*	P02351	26.61	6.69	36;30;28	14
Exponential	10	*rpsO*	P02371	10.14	10.4		2
Exponential	10	*rpsQ*	P02373	9.57	9.64		2
Exponential	10	*spoT*	P17580	79.34	8.89	40	
Exponential	10	*tgt*	P19675	42.59	5.97	17	
Exponential	10	*ttk*	P06969	22.84	8.78	22	
Exponential	10	*tufA*	P02990	43.18	5.3	47	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	10	*yaaA*	P11288	29.59	6.86	30	
Exponential	10	*yaeH*	P37048	15.1	6.61		2
Exponential	10	*ydjA*	P24250	20.06	6.31		6
Exponential	10	*yfiF*	P33635	37.78	8.94	40	
Exponential	10	*yihA*	P24253	23.56	6.86	26	
Exponential	10	*ykgM*	P71302	9.92	9.3		3
Exponential	11	*atpD*	P00824	50.19	4.9	45	
Exponential	11	*crp*	P03020	23.64	8.38	26	
Exponential	11	*fabR*	P27307	24.4	9.33	26	
Exponential	11	*glf*	P37747	42.97	6.62	30	
Exponential	11	*lon*	P08177	87.44	6.01	80	9
Exponential	11	*malP*	P00490	90.39	6.93	66	5
Exponential	11	*metE*	P25665	84.54	5.61	80;30	
Exponential	11	*oppA*	P23843	58.36	5.85	40	
Exponential	11	*parC*	P20082	83.83	6.24	66	
Exponential	11	*pheT*	P07395	87.38	5.17	66	
Exponential	11	*rhlB*	P24229	46.99	7.28	45	
Exponential	11	*rne*	P21513	118.2	5.48	80	
Exponential	11	*rob*	P27292	33.14	6.66	35	
Exponential	11	*rplB*	P02387	29.73	10.93	28	14
Exponential	11	*rplD*	P02388	22.09	9.72	16	5
Exponential	11	*rplE*	P02389	20.17	9.49	22	9
Exponential	11	*rplI*	P02418	15.77	6.17	17	7
Exponential	11	*rplM*	P02410	16.02	9.91	20	6
Exponential	11	*rplN*	P02411	13.54	10.43		4
Exponential	11	*rplO*	P02413	14.98	11.18		6
Exponential	11	*rplR*	P02419	12.77	10.42		3
Exponential	11	*rplS*	P02420	13	10.62	17;16	4
Exponential	11	*rplX*	P02425	11.19	10.21		5
Exponential	11	*rpmB*	P02428	8.88	11.42		6
Exponential	11	*rpmC*	P02429	7.27	9.98		3
Exponential	11	*rpoA*	P00574	36.51	4.98	35	6
Exponential	11	*rpoB*	P00575	150.63	5.15	140	15
Exponential	11	*rpoC*	P00577	155.16	6.67	140	17
Exponential	11	*rpsB*	P02351	26.61	6.69	30	7
Exponential	11	*rpsE*	P02356	17.47	10.11	20	4
Exponential	11	*rpsO*	P02371	10.14	10.4		4
Exponential	11	*rpsP*	P02372	9.19	10.54		5
Exponential	11	*rpsQ*	P02373	9.57	9.64		2
Exponential	11	*rpsS*	P02375	10.3	10.52		5
Exponential	11	*rsmC*	P39406	37.49	6	35	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	11	*ydjA*	P24250	20.06	6.31		7
Exponential	11	*yejK*	P33920	37.69	4.88	35	
Exponential	11	*yfhD*	P30135	53.16	5.03		2
Exponential	11	*yfiF*	P33635	37.78	8.94	35	
Exponential	11	*yggH*	P32049	27.31	6.42		3
Exponential	11	*ykgM*	P71302	9.92	9.3		1
Exponential	12	*atpD*	P00824	50.19	4.9	44	
Exponential	12	*hupA*	P02342	9.53	9.57		4
Exponential	12	*infC*	P02999	20.56	9.54		5
Exponential	12	*metE*	P25665	84.54	5.61	75	
Exponential	12	*mraW*	P18595	34.88	6.07		4
Exponential	12	*rhlB*	P24229	46.99	7.28	44	
Exponential	12	*rne*	P21513	118.2	5.48	130;75	
Exponential	12	*rnr*	P21499	92.11	8.78	75	
Exponential	12	*rplB*	P02387	29.73	10.93	30;27	10
Exponential	12	*rplD*	P02388	22.09	9.72	25	4
Exponential	12	*rplM*	P02410	16.02	9.91	25;19	8
Exponential	12	*rplO*	P02413	14.98	11.18		4
Exponential	12	*rplR*	P02419	12.77	10.42		3
Exponential	12	*rplS*	P02420	13	10.62		2
Exponential	12	*rplX*	P02425	11.19	10.21		4
Exponential	12	*rpmB*	P02428	8.88	11.42		4
Exponential	12	*rpmC*	P02429	7.27	9.98		2
Exponential	12	*rpoA*	P00574	36.51	4.98	40	
Exponential	12	*rpoB*	P00575	150.63	5.15	130	
Exponential	12	*rpoC*	P00577	155.16	6.67	130	
Exponential	12	*rpsC*	P02352	25.85	10.27	27	10
Exponential	12	*rpsD*	P02354	23.34	10.05		3
Exponential	12	*rpsE*	P02356	17.47	10.11	19	6
Exponential	12	*rpsG*	P02359	19.89	10.37		2
Exponential	12	*rpsI*	P02363	14.73	10.94		3
Exponential	12	*rpsM*	P02369	12.97	10.78		4
Exponential	12	*rpsP*	P02372	9.19	10.54		3
Exponential	12	*rpsS*	P02375	10.3	10.52		4
Exponential	12	*rsmC*	P39406	37.49	6	40	2
Exponential	12	*rsuA*	P33918	25.87	5.75	30	
Exponential	12	*selB*	P14081	68.87	6.11	48	
Exponential	12	*yhhX*	P46853	38.77	6.07	46	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	13	*deaD*	P23304	70.41	8.76	55;47	8
Exponential	13	*hupA*	P02342	9.53	9.57		3
Exponential	13	*infC*	P02999	20.56	9.54	21	6
Exponential	13	*kdgR*	P76268	30.03	5.43	30	
Exponential	13	*metE*	P25665	84.54	5.61	70	
Exponential	13	*panC*	P31663	31.6	5.92	70	
Exponential	13	*rhlE*	P25888	49.99	10.06	45	
Exponential	13	*rnr*	P21499	92.11	8.78	70	
Exponential	13	*rplB*	P02387	29.73	10.93	30;27	9
Exponential	13	*rplD*	P02388	22.09	9.72	125;26	6
Exponential	13	*rplE*	P02389	20.17	9.49	47	
Exponential	13	*rplM*	P02410	16.02	9.91		8
Exponential	13	*rplO*	P02413	14.98	11.18		5
Exponential	13	*rplP*	P02414	15.28	11.22		3
Exponential	13	*rplQ*	P02416	14.36	11.05		2
Exponential	13	*rplR*	P02419	12.77	10.42		3
Exponential	13	*rplU*	P02422	11.56	9.85		2
Exponential	13	*rplW*	P02424	11.2	9.94		4
Exponential	13	*rpmB*	P02428	8.88	11.42		5
Exponential	13	*rpoA*	P00574	36.51	4.98	40	
Exponential	13	*rpsC*	P02352	25.85	10.27	125;70;55;27	11
Exponential	13	*rpsD*	P02354	23.34	10.05	125;70;55;26;25	8
Exponential	13	*rpsE*	P02356	17.47	10.11	18	9
Exponential	13	*rpsG*	P02359	19.89	10.37	40;21	6
Exponential	13	*rpsI*	P02363	14.73	10.94		3
Exponential	13	*rpsM*	P02369	12.97	10.78		6
Exponential	13	*rpsS*	P02375	10.3	10.52	15	6
Exponential	13	*rpsT*	P02378	9.55	11.18		3
Exponential	13	*yafL*	Q47151	27.02	9.59		2
Exponential	14	*hupA*	P02342	9.53	9.57		5
Exponential	14	*hupB*	P02341	9.23	9.69		5
Exponential	14	*rplB*	P02387	29.73	10.93	30	11
Exponential	14	*rplM*	P02410	16.02	9.91		2
Exponential	14	*rplO*	P02413	14.98	11.18	40	7
Exponential	14	*rplP*	P02414	15.28	11.22	16	4
Exponential	14	*rplQ*	P02416	14.36	11.05	40;16	4
Exponential	14	*rplR*	P02419	12.77	10.42		5
Exponential	14	*rplU*	P02422	11.56	9.85		2
Exponential	14	*rpmB*	P02428	8.88	11.42		4

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	14	*rpoA*	P00574	36.51	4.98	40	
Exponential	14	*rpsC*	P02352	25.85	10.27	125;47;40	13
Exponential	14	*rpsD*	P02354	23.34	10.05	125;40;24	15
Exponential	14	*rpsE*	P02356	17.47	10.11	40;16	9
Exponential	14	*rpsF*	P02358	15.32	5.16		4
Exponential	14	*rpsG*	P02359	19.89	10.37	40;20;16	8
Exponential	14	*rpsI*	P02363	14.73	10.94		3
Exponential	14	*rpsM*	P02369	12.97	10.78		7
Exponential	14	*rpsR*	P02374	8.86	10.6		2
Exponential	14	*rpsS*	P02375	10.3	10.52		3
Exponential	14	*rpsT*	P02378	9.55	11.18		2
Exponential	15	*hupA*	P02342	9.53	9.57		4
Exponential	15	*hupB*	P02341	9.23	9.69		3
Exponential	15	*metE*	P25665	84.54	5.61	47	
Exponential	15	*pcnB*	P13685	54.66	9.68		6
Exponential	15	*rluC*	P23851	36.03	9.85		5
Exponential	15	*rplB*	P02387	29.73	10.93	30	5
Exponential	15	*rplE*	P02389	20.17	9.49		3
Exponential	15	*rplM*	P02410	16.02	9.91		3
Exponential	15	*rplO*	P02413	14.98	11.18	55	6
Exponential	15	*rplP*	P02414	15.28	11.22		5
Exponential	15	*rplQ*	P02416	14.36	11.05		5
Exponential	15	*rplS*	P02420	13	10.62		3
Exponential	15	*rplT*	P02421	13.37	11.47		2
Exponential	15	*rplU*	P02422	11.56	9.85		3
Exponential	15	*rplV*	P02423	12.23	10.23		5
Exponential	15	*rpmB*	P02428	8.88	11.42		3
Exponential	15	*rpoC*	P00577	155.16	6.67	55	
Exponential	15	*rpsC*	P02352	25.85	10.27	55;47	10
Exponential	15	*rpsD*	P02354	23.34	10.05	55;24	14
Exponential	15	*rpsE*	P02356	17.47	10.11	17	6
Exponential	15	*rpsF*	P02358	15.32	5.16		8
Exponential	15	*rpsG*	P02359	19.89	10.37	34;19;17	10
Exponential	15	*rpsI*	P02363	14.73	10.94		5
Exponential	15	*rpsK*	P02366	13.71	11.33		6
Exponential	15	*rpsM*	P02369	12.97	10.78		7
Exponential	15	*rpsR*	P02374	8.86	10.6		2
Exponential	15	*rpsS*	P02375	10.3	10.52		1
Exponential	15	*rpsT*	P02378	9.55	11.18		4

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Exponential	16	*hupA*	P02342	9.53	9.57		3
Exponential	16	*hupB*	P02341	9.23	9.69		2
Exponential	16	*hycE*	P16431	61.15	5.66	33	
Exponential	16	*ihfA*	P06984	11.35	9.34		2
Exponential	16	*ihfB*	P08756	10.65	9.34		2
Exponential	16	*metE*	P25665	84.54	5.61	70	
Exponential	16	*pcnB*	P13685	54.66	9.68	40	5
Exponential	16	*pgk*	P11665	40.99	5.08	125	
Exponential	16	*rplB*	P02387	29.73	10.93	29	5
Exponential	16	*rplE*	P02389	20.17	9.49		2
Exponential	16	*rplM*	P02410	16.02	9.91		1
Exponential	16	*rplO*	P02413	14.98	11.18		4
Exponential	16	*rplP*	P02414	15.28	11.22		6
Exponential	16	*rplQ*	P02416	14.36	11.05	70;19	4
Exponential	16	*rplT*	P02421	13.37	11.47		2
Exponential	16	*rplU*	P02422	11.56	9.85		2
Exponential	16	*rplV*	P02423	12.23	10.23		8
Exponential	16	*rpmB*	P02428	8.88	11.42		2
Exponential	16	*rpsB*	P02351	26.61	6.69	40	
Exponential	16	*rpsC*	P02352	25.85	10.27	70;25	9
Exponential	16	*rpsD*	P02354	23.34	10.05	25	6
Exponential	16	*rpsE*	P02356	17.47	10.11		2
Exponential	16	*rpsF*	P02358	15.32	5.16	19	8
Exponential	16	*rpsG*	P02359	19.89	10.37	55;29;24;19;17	9
Exponential	16	*rpsI*	P02363	14.73	10.94	125;70;55;19	7
Exponential	16	*rpsJ*	P02364	11.74	9.68		2
Exponential	16	*rpsK*	P02366	13.71	11.33		2
Exponential	16	*rpsM*	P02369	12.97	10.78		7
Exponential	16	*rpsR*	P02374	8.86	10.6	70;19	2
Exponential	16	*rpsT*	P02378	9.55	11.18		4
Exponential	16	*ybiB*	P30177	35.05	6.38	29	2

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	1	*accC*	P24182	49.32	6.65		7
Stationary	1	*adhE*	P17547	96	6.33	120;65	
Stationary	1	*atpF*	P00859	17.26	5.99	25	
Stationary	1	*dksA*	P18274	17.53	5.06	25	
Stationary	1	*fabB*	P14926	42.61	5.35		6
Stationary	1	*fabH*	P24249	33.52	5.08		5
Stationary	1	*fabI*	P29132	27.73	5.58		5
Stationary	1	*ghrA*	P75913	36.84	6.76	40	
Stationary	1	*glnS*	P00962	63.35	5.89		3
Stationary	1	*gltB*	P09831	161.93	6.19	175;150;120;100;65	
Stationary	1	*gltX*	P04805	53.82	5.59		6
Stationary	1	*gnd*	P00350	51.48	5.05	53;44;33;31;28;25	11
Stationary	1	*guaB*	P06981	52.02	6.02	65;62	
Stationary	1	*hns*	P08936	15.41	5.44		2
Stationary	1	*infB*	P02995	97.35	5.8	78	
Stationary	1	*oppA*	P23843	58.36	5.85	65	
Stationary	1	*pdxH*	P28225	25.41	9.18	31	
Stationary	1	*pgk*	P11665	40.99	5.08	48;45;37;31	22
Stationary	1	*rplC*	P02386	22.24	9.9		4
Stationary	1	*rplJ*	P02408	17.58	9.04		2
Stationary	1	*rplL*	P02392	12.16	4.6		4
Stationary	1	*rplS*	P02420	13	10.62		3
Stationary	1	*skp*	P11457	15.69	9.52	28;25	
Stationary	1	*sucB*	P07016	43.88	5.58	58	
Stationary	1	*thrS*	P00955	74.01	5.8	78	
Stationary	1	*trpS*	P00954	37.44	6.27	44	2
Stationary	1	*yggS*	P52054	25.79	6.09	31	
Stationary	1	*yicC*	P23839	33.17	5.1	40	
Stationary	1	*ykgM*	P71302	9.92	9.3		3
Stationary	1	*ynhG*	P76193	33.67	9.21	28	
Stationary	1	*zwf*	P22992	55.7	5.56	58	
Stationary	2	*accC*	P24182	49.32	6.65		13
Stationary	2	*aceE*	P06958	99.54	5.46		9
Stationary	2	*adhE*	P17547	96	6.33	155;125;63	13
Stationary	2	*arcA*	P03026	27.29	5.21		1
Stationary	2	*def*	P27251	19.2	5.23		5
Stationary	2	*dksA*	P18274	17.53	5.06		2
Stationary	2	*fabB*	P14926	42.61	5.35	54	11
Stationary	2	*fabG*	P25716	25.56	6.76		5

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	2	*fabH*	P24249	33.52	5.08	40	9
Stationary	2	*fabI*	P29132	27.73	5.58		7
Stationary	2	*glgB*	P07762	84.34	5.91	107	3
Stationary	2	*glnS*	P00962	63.35	5.89	75;30	12
Stationary	2	*gltB*	P09831	161.93	6.19	180;155;125	22
Stationary	2	*gltX*	P04805	53.82	5.59		14
Stationary	2	*gnd*	P00350	51.48	5.05	54	16
Stationary	2	*guaB*	P06981	52.02	6.02	63	6
Stationary	2	*hns*	P08936	15.41	5.44		3
Stationary	2	*infB*	P02995	97.35	5.8	85	6
Stationary	2	*infC*	P02999	20.56	9.54		1
Stationary	2	*lpd*	P00391	50.56	5.79		8
Stationary	2	*map*	P07906	29.33	5.64		1
Stationary	2	*mfd*	P30958	129.98	5.79	125	20
Stationary	2	*mprA*	P24201	20.56	5.82	28;23	3
Stationary	2	*mukB*	P22523	170.23	5.24		5
Stationary	2	*nusG*	P16921	20.4	6.33		2
Stationary	2	*ompR*	P03025	27.35	6.04		8
Stationary	2	*oppA*	P23843	58.36	5.85		6
Stationary	2	*pdxH*	P28225	25.41	9.18		4
Stationary	2	*pgk*	P11665	40.99	5.08	125;47;23	29
Stationary	2	*purR*	P15039	38.04	6.3		4
Stationary	2	*rbfA*	P09170	15.02	5.96		3
Stationary	2	*rplA*	P02384	24.6	9.64		2
Stationary	2	*rplB*	P02387	29.73	10.93		2
Stationary	2	*rplC*	P02386	22.24	9.9		4
Stationary	2	*rplD*	P02388	22.09	9.72		4
Stationary	2	*rplE*	P02389	20.17	9.49		7
Stationary	2	*rplI*	P02418	15.77	6.17		3
Stationary	2	*rplJ*	P02408	17.58	9.04		3
Stationary	2	*rplK*	P02409	14.74	9.64		1
Stationary	2	*rplL*	P02392	12.16	4.6		5
Stationary	2	*rplM*	P02410	16.02	9.91		2
Stationary	2	*rplS*	P02420	13	10.62		4
Stationary	2	*rplU*	P02422	11.56	9.85		2
Stationary	2	*rplX*	P02425	11.19	10.21		1
Stationary	2	*rplY*	P02426	10.69	9.6		5
Stationary	2	*rpmA*	P02427	8.99	10.58		2
Stationary	2	*rpmB*	P02428	8.88	11.42		1
Stationary	2	*rpmC*	P02429	7.27	9.98		1

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	2	*rpmD*	P02430	6.41	10.96		2
Stationary	2	*rpoA*	P00574	36.51	4.98		6
Stationary	2	*rpsA*	P02349	61.16	4.89	85	8
Stationary	2	*rpsH*	P02361	14	9.44		3
Stationary	2	*rpsJ*	P02364	11.74	9.68		1
Stationary	2	*rpsM*	P02369	12.97	10.78		3
Stationary	2	*rpsP*	P02372	9.19	10.54		2
Stationary	2	*rpsR*	P02374	8.86	10.6		2
Stationary	2	*rpsT*	P02378	9.55	11.18		3
Stationary	2	*skp*	P11457	15.69	9.52	23	1
Stationary	2	*slyA*	P55740	16.35	6.6		3
Stationary	2	*sucA*	P07015	105.06	6.04	107	9
Stationary	2	*sucB*	P07016	43.88	5.58	58	8
Stationary	2	*thrS*	P00955	74.01	5.8	85	7
Stationary	2	*tig*	P22257	48.19	4.83		6
Stationary	2	*trpS*	P00954	37.44	6.27	44	8
Stationary	2	*tufA*	P02990	43.18	5.3		8
Stationary	2	*ycgK*	P76002	12.52	9.01		5
Stationary	2	*yciK*	P31808	27.93	7.67		4
Stationary	2	*yggS*	P52054	25.79	6.09	30	
Stationary	2	*yhaJ*	P42623	33.26	6.05		4
Stationary	2	*yicC*	P23839	33.17	5.1		3
Stationary	2	*ykgM*	P71302	9.92	9.3		6
Stationary	2	*ynhG*	P76193	33.67	9.21		4
Stationary	2	*zwf*	P22992	55.7	5.56	58	4
Stationary	3	*accC*	P24182	49.32	6.65	58	6
Stationary	3	*adhE*	P17547	96	6.33	125	12
Stationary	3	*fabB*	P14926	42.61	5.35		4
Stationary	3	*fabH*	P24249	33.52	5.08		6
Stationary	3	*fabI*	P29132	27.73	5.58		3
Stationary	3	*glgB*	P07762	84.34	5.91		2
Stationary	3	*glnS*	P00962	63.35	5.89	75	3
Stationary	3	*gltB*	P09831	161.93	6.19	180	7
Stationary	3	*gltX*	P04805	53.82	5.59	63	4
Stationary	3	*gnd*	P00350	51.48	5.05	58;54	12
Stationary	3	*guaB*	P06981	52.02	6.02		5
Stationary	3	*infB*	P02995	97.35	5.8		3
Stationary	3	*lpd*	P00391	50.56	5.79		7
Stationary	3	*mprA*	P24201	20.56	5.82	24	1
Stationary	3	*ompR*	P03025	27.35	6.04	35	4

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	3	*pgk*	P11665	40.99	5.08	150;80;47;44;40;35;31;27	14
Stationary	3	*phoB*	P08402	26.43	5.46		2
Stationary	3	*rbfA*	P09170	15.02	5.96		4
Stationary	3	*rnr*	P21499	92.11	8.78	80	
Stationary	3	*rplA*	P02384	24.6	9.64		1
Stationary	3	*rplB*	P02387	29.73	10.93		2
Stationary	3	*rplI*	P02418	15.77	6.17		2
Stationary	3	*rplJ*	P02408	17.58	9.04		2
Stationary	3	*rplL*	P02392	12.16	4.6		4
Stationary	3	*rplS*	P02420	13	10.62		2
Stationary	3	*rplU*	P02422	11.56	9.85		2
Stationary	3	*rplY*	P02426	10.69	9.6		4
Stationary	3	*rpoA*	P00574	36.51	4.98		3
Stationary	3	*rpsA*	P02349	61.16	4.89		3
Stationary	3	*skp*	P11457	15.69	9.52		1
Stationary	3	*slyA*	P55740	16.35	6.6		2
Stationary	3	*sucA*	P07015	105.06	6.04		11
Stationary	3	*sucB*	P07016	43.88	5.58		7
Stationary	3	*thrS*	P00955	74.01	5.8		6
Stationary	3	*trpS*	P00954	37.44	6.27		2
Stationary	3	*ycgK*	P76002	12.52	9.01		4
Stationary	3	*yciK*	P31808	27.93	7.67		1
Stationary	3	*yicC*	P23839	33.17	5.1		4
Stationary	3	*ykgM*	P71302	9.92	9.3		1
Stationary	3	*ynhG*	P76193	33.67	9.21		3
Stationary	3	*zwf*	P22992	55.7	5.56		2
Stationary	4	*accA*	P30867	35.11	5.76		6
Stationary	4	*accC*	P24182	49.32	6.65	57	
Stationary	4	*accD*	P08193	33.32	7.58		2
Stationary	4	*aceE*	P06958	99.54	5.46		4
Stationary	4	*adhE*	P17547	96	6.33	120;100	11
Stationary	4	*dksA*	P18274	17.53	5.06		1
Stationary	4	*fabB*	P14926	42.61	5.35	54	2
Stationary	4	*fabG*	P25716	25.56	6.76		3
Stationary	4	*fabH*	P24249	33.52	5.08		3
Stationary	4	*ghrA*	P75913	36.84	6.76		4
Stationary	4	*glnS*	P00962	63.35	5.89	80	5

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	4	*gltB*	P09831	161.93	6.19	170;145;100;68	32
Stationary	4	*gltX*	P04805	53.82	5.59	63	
Stationary	4	*gnd*	P00350	51.48	5.05	54	21
Stationary	4	*guaB*	P06981	52.02	6.02		16
Stationary	4	*hns*	P08936	15.41	5.44		3
Stationary	4	*infB*	P02995	97.35	5.8		8
Stationary	4	*lpd*	P00391	50.56	5.79		4
Stationary	4	*mfd*	P30958	129.98	5.79	145	15
Stationary	4	*nikE*	P33594	29.72	10.11	47	
Stationary	4	*oppA*	P23843	58.36	5.85	68	
Stationary	4	*pgk*	P11665	40.99	5.08	145;47;44;40;35;31;28;25;22	15
Stationary	4	*rbfA*	P09170	15.02	5.96		6
Stationary	4	*rplB*	P02387	29.73	10.93		4
Stationary	4	*rplC*	P02386	22.24	9.9		3
Stationary	4	*rplI*	P02418	15.77	6.17		1
Stationary	4	*rplJ*	P02408	17.58	9.04		1
Stationary	4	*rplL*	P02392	12.16	4.6		4
Stationary	4	*rplS*	P02420	13	10.62		5
Stationary	4	*rplY*	P02426	10.69	9.6		3
Stationary	4	*skp*	P11457	15.69	9.52		4
Stationary	4	*sucA*	P07015	105.06	6.04		3
Stationary	4	*sucB*	P07016	43.88	5.58		6
Stationary	4	*thrS*	P00955	74.01	5.8		15
Stationary	4	*trpS*	P00954	37.44	6.27		9
Stationary	4	*ycgK*	P76002	12.52	9.01		8
Stationary	4	*yegQ*	P76403	51.19	5.8		4
Stationary	4	*ykgM*	P71302	9.92	9.3		1
Stationary	4	*ynhG*	P76193	33.67	9.21		5
Stationary	4	*zwf*	P22992	55.7	5.56		9
Stationary	5	*accA*	P30867	35.11	5.76		8
Stationary	5	*accD*	P08193	33.32	7.58		6
Stationary	5	*aceA*	P05313	47.52	5.16	47	
Stationary	5	*adhE*	P17547	96	6.33		2
Stationary	5	*atpF*	P00859	17.26	5.99	26	
Stationary	5	*bfr*	P11056	18.5	4.69	23	
Stationary	5	*dksA*	P18274	17.53	5.06		3
Stationary	5	*erfK*	P39176	32.08	7.97		2
Stationary	5	*ghrA*	P75913	36.84	6.76		3
Stationary	5	*gltB*	P09831	161.93	6.19		28

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	5	*gnd*	P00350	51.48	5.05		21
Stationary	5	*guaB*	P06981	52.02	6.02		7
Stationary	5	*hns*	P08936	15.41	5.44		5
Stationary	5	*infB*	P02995	97.35	5.8		6
Stationary	5	*manX*	P08186	34.92	5.74		4
Stationary	5	*pgk*	P11665	40.99	5.08		7
Stationary	5	*purU*	P37051	31.93	6.5		5
Stationary	5	*rbfA*	P09170	15.02	5.96		2
Stationary	5	*rplK*	P02409	14.74	9.64		1
Stationary	5	*rplS*	P02420	13	10.62		6
Stationary	5	*rplY*	P02426	10.69	9.6		1
Stationary	5	*skp*	P11457	15.69	9.52		4
Stationary	5	*thrS*	P00955	74.01	5.8		9
Stationary	5	*trpS*	P00954	37.44	6.27		4
Stationary	5	*tyrS*	P00951	47.4	5.59		7
Stationary	5	*ybaK*	P37175	17.09	9.02		3
Stationary	5	*ycgK*	P76002	12.52	9.01		7
Stationary	5	*ydgH*	P76177	31.91	9.1		8
Stationary	5	*ykgM*	P71302	9.92	9.3		2
Stationary	5	*ynhG*	P76193	33.67	9.21		2
Stationary	5	*zwf*	P22992	55.7	5.56		2
Stationary	6	*accA*	P30867	35.11	5.76		7
Stationary	6	*accD*	P08193	33.32	7.58		5
Stationary	6	*aceA*	P05313	47.52	5.16	55	
Stationary	6	*aceE*	P06958	99.54	5.46	107	9
Stationary	6	*aceF*	P06959	65.96	5.09	95	6
Stationary	6	*adhE*	P17547	96	6.33	170;107	
Stationary	6	*ahpC*	P26427	20.63	5.03	29	
Stationary	6	*arcA*	P03026	27.29	5.21		8
Stationary	6	*aroC*	P12008	39.01	5.82		2
Stationary	6	*atpD*	P00824	50.19	4.9		2
Stationary	6	*carB*	P00968	117.71	5.22	170	
Stationary	6	*cbpA*	P36659	34.46	6.33		4
Stationary	6	*cpsB*	P24174	53.02	5.2		4
Stationary	6	*crp*	P03020	23.64	8.38	30	
Stationary	6	*dacB*	P24228	49.57	8.73		4
Stationary	6	*dinB*	Q47155	39.52	9.28		3
Stationary	6	*dksA*	P18274	17.53	5.06		4
Stationary	6	*dnaK*	P04475	68.98	4.83	75	
Stationary	6	*dps*	P27430	18.56	5.72		5
Stationary	6	*eda*	P10177	22.28	5.57	30	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	6	*fabA*	P18391	18.84	6.16		8
Stationary	6	*fabG*	P25716	25.56	6.76		6
Stationary	6	*fmt*	P23882	34.04	5.56		8
Stationary	6	*ftnA*	P23887	19.42	4.77	23	
Stationary	6	*glgB*	P07762	84.34	5.91		9
Stationary	6	*glnK*	P38504	12.26	5.84		3
Stationary	6	*gltB*	P09831	161.93	6.19	170	18
Stationary	6	*gnd*	P00350	51.48	5.05		12
Stationary	6	*gpmA*	P31217	28.43	5.86	34	
Stationary	6	*groL*	P06139	57.2	4.85	65;60	
Stationary	6	*grxB*	P39811	24.35	7.72	32	
Stationary	6	*guaB*	P06981	52.02	6.02		4
Stationary	6	*gyrA*	P09097	96.96	5.09		20
Stationary	6	*hisS*	P04804	46.9	5.65		6
Stationary	6	*hns*	P08936	15.41	5.44		8
Stationary	6	*htpG*	P10413	71.42	5.09	75	
Stationary	6	*icd*	P08200	45.76	5.15	55	
Stationary	6	*ilvE*	P00510	33.96	5.54		2
Stationary	6	*infB*	P02995	97.35	5.8		12
Stationary	6	*iscS*	P39171	45.09	5.94		9
Stationary	6	*ivy*	P45502	14.1	5.51		5
Stationary	6	*ligA*	P15042	73.61	5.39		12
Stationary	6	*lrp*	P19494	18.76	8.91		5
Stationary	6	*manX*	P08186	34.92	5.74		10
Stationary	6	*map*	P07906	29.33	5.64		10
Stationary	6	*mdh*	P06994	32.34	5.61	39	
Stationary	6	*mdoG*	P33136	55.37	6.26		20
Stationary	6	*metE*	P25665	84.54	5.61		4
Stationary	6	*metJ*	P08338	12.01	5.4		5
Stationary	6	*metQ*	P28635	27.24	4.93	34	
Stationary	6	*nadR*	P27278	47.35	5.42		8
Stationary	6	*nagA*	P15300	40.95	5.65		2
Stationary	6	*ompC*	P06996	38.31	4.48	45	
Stationary	6	*pdxH*	P28225	25.41	9.18		6
Stationary	6	*pepA*	P11648	54.88	6.82		5
Stationary	6	*pgk*	P11665	40.99	5.08		8
Stationary	6	*polA*	P00582	103.12	5.4		8
Stationary	6	*ppc*	P00864	99.06	5.52	95	
Stationary	6	*prc*	P23865	74.32	6.04		11
Stationary	6	*purC*	P21155	27	5.07	34	
Stationary	6	*purU*	P37051	31.93	6.5		6

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	6	*rbsA*	P04983	55.04	5.84	36	
Stationary	6	*relE*	P07008	11.23	9.67		2
Stationary	6	*rho*	P03002	47	6.75		7
Stationary	6	*rnc*	P05797	25.55	6.4		6
Stationary	6	*rplA*	P02384	24.6	9.64	36	5
Stationary	6	*rplB*	P02387	29.73	10.93		4
Stationary	6	*rplC*	P02386	22.24	9.9		5
Stationary	6	*rplD*	P02388	22.09	9.72	30	
Stationary	6	*rplE*	P02389	20.17	9.49	27	6
Stationary	6	*rplF*	P02390	18.77	9.71	75	5
Stationary	6	*rplI*	P02418	15.77	6.17	23	7
Stationary	6	*rplJ*	P02408	17.58	9.04		3
Stationary	6	*rplK*	P02409	14.74	9.64		6
Stationary	6	*rplL*	P02392	12.16	4.6		2
Stationary	6	*rplO*	P02413	14.98	11.18		5
Stationary	6	*rplQ*	P02416	14.36	11.05	25;23	
Stationary	6	*rplS*	P02420	13	10.62		5
Stationary	6	*rplX*	P02425	11.19	10.21		3
Stationary	6	*rplY*	P02426	10.69	9.6		3
Stationary	6	*rpmA*	P02427	8.99	10.58		2
Stationary	6	*rpmD*	P02430	6.41	10.96		3
Stationary	6	*rpoA*	P00574	36.51	4.98		7
Stationary	6	*rpoB*	P00575	150.63	5.15		6
Stationary	6	*rpsA*	P02349	61.16	4.89	75	
Stationary	6	*rpsB*	P02351	26.61	6.69		6
Stationary	6	*rpsD*	P02354	23.34	10.05	32	
Stationary	6	*rpsE*	P02356	17.47	10.11	25	
Stationary	6	*rpsI*	P02363	14.73	10.94	25	
Stationary	6	*rpsJ*	P02364	11.74	9.68		3
Stationary	6	*rpsM*	P02369	12.97	10.78		2
Stationary	6	*rpsS*	P02375	10.3	10.52		2
Stationary	6	*rpsT*	P02378	9.55	11.18		2
Stationary	6	*rpsU*	P02379	8.37	11.15		4
Stationary	6	*selA*	P23328	50.61	6.05		6
Stationary	6	*serA*	P08328	44.04	5.93	55	
Stationary	6	*skp*	P11457	15.69	9.52	25	5
Stationary	6	*sodB*	P09157	21.13	5.58	30;29	
Stationary	6	*sthA*	P27306	51.43	6.09		2
Stationary	6	*stpA*	P30017	15.35	7.95		6
Stationary	6	*talB*	P30148	35.09	5.11	43	
Stationary	6	*trpS*	P00954	37.44	6.27		2
Stationary	6	*truA*	P07649	30.4	8.68		4

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	6	*tufA*	P02990	43.18	5.3	51	
Stationary	6	*tyrS*	P00951	47.4	5.59		7
Stationary	6	*uvrB*	P07025	76.09	5.14		4
Stationary	6	*wrbA*	P30849	20.71	5.6	30	
Stationary	6	*xthA*	P09030	30.97	5.8		6
Stationary	6	*yaeB*	P28634	26.36	7.86		3
Stationary	6	*yaeH*	P37048	15.1	6.61		2
Stationary	6	*yaiN*	P55756	10.32	5.84		2
Stationary	6	*ybeX*	P77392	33.3	4.55		3
Stationary	6	*ybeZ*	P77349	40.65	6.24		7
Stationary	6	*ycgK*	P76002	12.52	9.01		5
Stationary	6	*yciH*	P08245	11.4	9.38		3
Stationary	6	*ydcP*	P76104	70.58	6.69	39	
Stationary	6	*ydgH*	P76177	31.91	9.1		14
Stationary	6	*yeaO*	P76243	13.39	6.29		7
Stationary	6	*yffB*	P24178	13.6	6.09		2
Stationary	6	*yggS*	P52054	25.79	6.09		3
Stationary	6	*yhbJ*	P33995	32.49	6.72		2
Stationary	6	*yhhK*	P37613	14.51	6.59		2
Stationary	6	*ykgM*	P71302	9.92	9.3		5
Stationary	6	*zwf*	P22992	55.7	5.56		3
Stationary	7	*aceA*	P05313	47.52	5.16	58	
Stationary	7	*adhE*	P17547	96	6.33	115	
Stationary	7	*ahpC*	P26427	20.63	5.03	30	
Stationary	7	*arcA*	P03026	27.29	5.21		7
Stationary	7	*aroC*	P12008	39.01	5.82		2
Stationary	7	*atpA*	P00822	55.22	5.8	65	
Stationary	7	*carB*	P00968	117.71	5.22	155	
Stationary	7	*dnaK*	P04475	68.98	4.83	78	
Stationary	7	*dnaX*	P06710	71.14	6.39		2
Stationary	7	*eda*	P10177	22.28	5.57	32	
Stationary	7	*fabG*	P25716	25.56	6.76		7
Stationary	7	*fmt*	P23882	34.04	5.56		4
Stationary	7	*gltB*	P09831	161.93	6.19	180	
Stationary	7	*groL*	P06139	57.2	4.85	65	
Stationary	7	*grxB*	P39811	24.35	7.72	32	
Stationary	7	*gyrA*	P09097	96.96	5.09		10
Stationary	7	*hisS*	P04804	46.9	5.65		16
Stationary	7	*hns*	P08936	15.41	5.44		4
Stationary	7	*infA*	P02998	8.12	9.23		4
Stationary	7	*kdgR*	P76268	30.03	5.43	30	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	7	*lacZ*	P00722	116.35	5.28	145;65	
Stationary	7	*map*	P07906	29.33	5.64		2
Stationary	7	*metJ*	P08338	12.01	5.4		6
Stationary	7	*moaB*	P30746	18.53	5.73		4
Stationary	7	*rho*	P03002	47	6.75		4
Stationary	7	*rlmB*	P39290	26.56	6.17		2
Stationary	7	*rplA*	P02384	24.6	9.64	40	1
Stationary	7	*rplB*	P02387	29.73	10.93		3
Stationary	7	*rplC*	P02386	22.24	9.9		4
Stationary	7	*rplE*	P02389	20.17	9.49	27	6
Stationary	7	*rplF*	P02390	18.77	9.71		7
Stationary	7	*rplI*	P02418	15.77	6.17		9
Stationary	7	*rplJ*	P02408	17.58	9.04		5
Stationary	7	*rplL*	P02392	12.16	4.6		2
Stationary	7	*rplM*	P02410	16.02	9.91		1
Stationary	7	*rplO*	P02413	14.98	11.18		3
Stationary	7	*rplS*	P02420	13	10.62		5
Stationary	7	*rplX*	P02425	11.19	10.21		6
Stationary	7	*rpmD*	P02430	6.41	10.96		3
Stationary	7	*rpsA*	P02349	61.16	4.89	78	
Stationary	7	*rpsB*	P02351	26.61	6.69	40	9
Stationary	7	*rpsD*	P02354	23.34	10.05	32	
Stationary	7	*rpsJ*	P02364	11.74	9.68		3
Stationary	7	*rpsQ*	P02373	9.57	9.64		2
Stationary	7	*rpsT*	P02378	9.55	11.18		2
Stationary	7	*serA*	P08328	44.04	5.93	53	
Stationary	7	*sodB*	P09157	21.13	5.58	30	
Stationary	7	*stpA*	P30017	15.35	7.95		5
Stationary	7	*suhB*	P22783	29.17	6.45		6
Stationary	7	*tpx*	P37901	17.7	4.75	27	
Stationary	7	*tsf*	P02997	30.29	5.22	40	
Stationary	7	*tufA*	P02990	43.18	5.3	53;43;40	
Stationary	7	*yjgA*	P26650	21.36	5.3		5
Stationary	7	*ykgM*	P71302	9.92	9.3		4
Stationary	8	*aceF*	P06959	65.96	5.09	75	
Stationary	8	*arcA*	P03026	27.29	5.21		1
Stationary	8	*crp*	P03020	23.64	8.38		3
Stationary	8	*fabG*	P25716	25.56	6.76		2
Stationary	8	*hflX*	P25519	48.33	5.68	54	
Stationary	8	*hisS*	P04804	46.9	5.65	54;52	10
Stationary	8	*hns*	P08936	15.41	5.44		1
Stationary	8	*infA*	P02998	8.12	9.23		2

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	9	*manX*	P08186	34.92	5.74	37	
Stationary	9	*moaB*	P30746	18.53	5.73	21	3
Stationary	9	*oxyR*	P11721	34.28	5.96	38	
Stationary	9	*parC*	P20082	83.83	6.24		5
Stationary	9	*pepA*	P11648	54.88	6.82	58	2
Stationary	9	*ppx*	P29014	58	6.69	58	
Stationary	9	*rho*	P03002	47	6.75	54	6
Stationary	9	*rplB*	P02387	29.73	10.93	30	12
Stationary	9	*rplC*	P02386	22.24	9.9	29	6
Stationary	9	*rplD*	P02388	22.09	9.72		6
Stationary	9	*rplE*	P02389	20.17	9.49	23	10
Stationary	9	*rplF*	P02390	18.77	9.71	24	9
Stationary	9	*rplI*	P02418	15.77	6.17		8
Stationary	9	*rplJ*	P02408	17.58	9.04		6
Stationary	9	*rplL*	P02392	12.16	4.6		4
Stationary	9	*rplM*	P02410	16.02	9.91		4
Stationary	9	*rplO*	P02413	14.98	11.18		5
Stationary	9	*rplR*	P02419	12.77	10.42		3
Stationary	9	*rplS*	P02420	13	10.62		6
Stationary	9	*rplX*	P02425	11.19	10.21		4
Stationary	9	*rplY*	P02426	10.69	9.6		2
Stationary	9	*rpmB*	P02428	8.88	11.42		3
Stationary	9	*rpmC*	P02429	7.27	9.98		2
Stationary	9	*rpoA*	P00574	36.51	4.98	45	10
Stationary	9	*rpoB*	P00575	150.63	5.15	175;140;110	19
Stationary	9	*rpoC*	P00577	155.16	6.67	175;140;110;63	15
Stationary	9	*rpsB*	P02351	26.61	6.69	37;33;32	13
Stationary	9	*rpsC*	P02352	25.85	10.27	30	5
Stationary	9	*rpsD*	P02354	23.34	10.05	29	
Stationary	9	*rpsE*	P02356	17.47	10.11	21	1
Stationary	9	*rpsJ*	P02364	11.74	9.68		2
Stationary	9	*rpsO*	P02371	10.14	10.4		3
Stationary	9	*rpsP*	P02372	9.19	10.54		5
Stationary	9	*rpsQ*	P02373	9.57	9.64		3
Stationary	9	*rpsR*	P02374	8.86	10.6		1
Stationary	9	*rpsS*	P02375	10.3	10.52		3
Stationary	9	*rpsT*	P02378	9.55	11.18		2
Stationary	9	*rsmC*	P39406	37.49	6	43	
Stationary	9	*secA*	P10408	102.02	5.43	110;105	5
Stationary	9	*stpA*	P30017	15.35	7.95		5
Stationary	9	*truD*	Q57261	39.09	6.15	43	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides #
Stationary	9	*tsf*	P02997	30.29	5.22		3
Stationary	9	*wcaI*	P32057	44.91	8.8		4
Stationary	9	*yaaA*	P11288	29.59	6.86		5
Stationary	9	*ydjA*	P24250	20.06	6.31	23	5
Stationary	9	*ygeA*	P03813	25.25	5.09	30	
Stationary	9	*yihA*	P24253	23.56	6.86	29	7
Stationary	9	*yihI*	P32130	19.06	6.37		2
Stationary	9	*ykgM*	P71302	9.92	9.3		3
Stationary	9	*yraL*	P45528	31.35	5.83		5
Stationary	10	*cbpA*	P36659	34.46	6.33	36;35;27	
Stationary	10	*clpA*	P15716	84.21	5.91		2
Stationary	10	*crp*	P03020	23.64	8.38	26	1
Stationary	10	*envC*	P37690	46.59	9.91	26	
Stationary	10	*hupA*	P02342	9.53	9.57		3
Stationary	10	*infC*	P02999	20.56	9.54	23	
Stationary	10	*lon*	P08177	87.44	6.01	110	9
Stationary	10	*malP*	P00490	90.39	6.93	100;66	9
Stationary	10	*malQ*	P15977	78.5	6.14	100	
Stationary	10	*mraW*	P18595	34.88	6.07		2
Stationary	10	*pgk*	P11665	40.99	5.08	36	
Stationary	10	*ppk*	P28688	80.3	8.96	110;100;66	3
Stationary	10	*ppx*	P29014	58	6.69	58	
Stationary	10	*rnr*	P21499	92.11	8.78		4
Stationary	10	*rob*	P27292	33.14	6.66	38	4
Stationary	10	*rplB*	P02387	29.73	10.93	35;33;29;27	12
Stationary	10	*rplD*	P02388	22.09	9.72	26;24	6
Stationary	10	*rplE*	P02389	20.17	9.49	24	8
Stationary	10	*rplI*	P02418	15.77	6.17		6
Stationary	10	*rplM*	P02410	16.02	9.91	21	7
Stationary	10	*rplO*	P02413	14.98	11.18		6
Stationary	10	*rplR*	P02419	12.77	10.42		3
Stationary	10	*rplS*	P02420	13	10.62		2
Stationary	10	*rplX*	P02425	11.19	10.21		6
Stationary	10	*rpmA*	P02427	8.99	10.58		2
Stationary	10	*rpmB*	P02428	8.88	11.42		6
Stationary	10	*rpmC*	P02429	7.27	9.98		3
Stationary	10	*rpoA*	P00574	36.51	4.98	49;35	12
Stationary	10	*rpoB*	P00575	150.63	5.15	140;110	24
Stationary	10	*rpoC*	P00577	155.16	6.67	175;140;110;100;66; 58;54	29
Stationary	10	*rpsB*	P02351	26.61	6.69	33	7
Stationary	10	*rpsC*	P02352	25.85	10.27	27	9

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	10	*rpsD*	P02354	23.34	10.05	27	4
Stationary	10	*rpsE*	P02356	17.47	10.11	21	4
Stationary	10	*rpsM*	P02369	12.97	10.78		3
Stationary	10	*rpsO*	P02371	10.14	10.4		3
Stationary	10	*rpsP*	P02372	9.19	10.54		5
Stationary	10	*rpsS*	P02375	10.3	10.52		4
Stationary	10	*rpsT*	P02378	9.55	11.18		4
Stationary	10	*rsmC*	P39406	37.49	6	45	
Stationary	10	*secA*	P10408	102.02	5.43	140	
Stationary	10	*speG*	P37354	21.76	6.22	25	2
Stationary	10	*spy*	P77754	15.88	9.45		6
Stationary	10	*ydjA*	P24250	20.06	6.31	24;21	8
Stationary	10	*yggH*	P32049	27.31	6.42		5
Stationary	11	*allR*	P77734	29.27	5.77		2
Stationary	11	*gadB*	P28302	52.67	5.29		2
Stationary	11	*hupA*	P02342	9.53	9.57		5
Stationary	11	*infC*	P02999	20.56	9.54	24	6
Stationary	11	*kdgR*	P76268	30.03	5.43	35	5
Stationary	11	*rnr*	P21499	92.11	8.78		8
Stationary	11	*rplB*	P02387	29.73	10.93	35;30	14
Stationary	11	*rplD*	P02388	22.09	9.72	27	6
Stationary	11	*rplE*	P02389	20.17	9.49		5
Stationary	11	*rplM*	P02410	16.02	9.91		10
Stationary	11	*rplN*	P02411	13.54	10.43		3
Stationary	11	*rplO*	P02413	14.98	11.18		7
Stationary	11	*rplP*	P02414	15.28	11.22		3
Stationary	11	*rplQ*	P02416	14.36	11.05		4
Stationary	11	*rplR*	P02419	12.77	10.42		6
Stationary	11	*rplS*	P02420	13	10.62		2
Stationary	11	*rplU*	P02422	11.56	9.85		2
Stationary	11	*rplV*	P02423	12.23	10.23		2
Stationary	11	*rplW*	P02424	11.2	9.94		3
Stationary	11	*rplX*	P02425	11.19	10.21		4
Stationary	11	*rpmA*	P02427	8.99	10.58		4
Stationary	11	*rpmB*	P02428	8.88	11.42		6
Stationary	11	*rpmC*	P02429	7.27	9.98		2
Stationary	11	*rpoA*	P00574	36.51	4.98		4
Stationary	11	*rpsC*	P02352	25.85	10.27	30;29	11
Stationary	11	*rpsD*	P02354	23.34	10.05		11
Stationary	11	*rpsE*	P02356	17.47	10.11		7
Stationary	11	*rpsF*	P02358	15.32	5.16		3
Stationary	11	*rpsG*	P02359	19.89	10.37		8

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	11	*rpsI*	P02363	14.73	10.94		4
Stationary	11	*rpsK*	P02366	13.71	11.33		6
Stationary	11	*rpsM*	P02369	12.97	10.78		6
Stationary	11	*rpsN*	P02370	11.45	11.16		3
Stationary	11	*rpsP*	P02372	9.19	10.54		5
Stationary	11	*rpsR*	P02374	8.86	10.6		2
Stationary	11	*rpsS*	P02375	10.3	10.52		4
Stationary	11	*rpsT*	P02378	9.55	11.18		5
Stationary	11	*yggH*	P32049	27.31	6.42		2
Stationary	11	*yhhX*	P46853	38.77	6.07		9
Stationary	12	*hupA*	P02342	9.53	9.57		4
Stationary	12	*hupB*	P02341	9.23	9.69		2
Stationary	12	*infC*	P02999	20.56	9.54	25;23	6
Stationary	12	*kdgR*	P76268	30.03	5.43	35;34	
Stationary	12	*relA*	P11585	83.88	6.29		2
Stationary	12	*rhlE*	P25888	49.99	10.06	54	
Stationary	12	*rplB*	P02387	29.73	10.93	35;34	10
Stationary	12	*rplD*	P02388	22.09	9.72	27;22	4
Stationary	12	*rplM*	P02410	16.02	9.91		5
Stationary	12	*rplO*	P02413	14.98	11.18		6
Stationary	12	*rplP*	P02414	15.28	11.22		4
Stationary	12	*rplQ*	P02416	14.36	11.05		3
Stationary	12	*rplR*	P02419	12.77	10.42		4
Stationary	12	*rplU*	P02422	11.56	9.85		2
Stationary	12	*rplV*	P02423	12.23	10.23		2
Stationary	12	*rpmB*	P02428	8.88	11.42		5
Stationary	12	*rpoC*	P00577	155.16	6.67	175	
Stationary	12	*rpsC*	P02352	25.85	10.27	175;105;65;54;46;40;35;34	10
Stationary	12	*rpsD*	P02354	23.34	10.05	27;25	10
Stationary	12	*rpsE*	P02356	17.47	10.11	54;20	8
Stationary	12	*rpsG*	P02359	19.89	10.37	23	5
Stationary	12	*rpsI*	P02363	14.73	10.94		3
Stationary	12	*rpsM*	P02369	12.97	10.78		4
Stationary	12	*rpsR*	P02374	8.86	10.6		2
Stationary	12	*rpsS*	P02375	10.3	10.52		6
Stationary	12	*rpsT*	P02378	9.55	11.18		3
Stationary	12	*yhhX*	P46853	38.77	6.07	46	

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides #
Stationary	13	*hupA*	P02342	9.53	9.57		8
Stationary	13	*hupB*	P02341	9.23	9.69		6
Stationary	13	*infC*	P02999	20.56	9.54		3
Stationary	13	*rluC*	P23851	36.03	9.85		4
Stationary	13	*rplB*	P02387	29.73	10.93		9
Stationary	13	*rplD*	P02388	22.09	9.72		5
Stationary	13	*rplM*	P02410	16.02	9.91		3
Stationary	13	*rplO*	P02413	14.98	11.18		11
Stationary	13	*rplP*	P02414	15.28	11.22		6
Stationary	13	*rplQ*	P02416	14.36	11.05		4
Stationary	13	*rplR*	P02419	12.77	10.42		3
Stationary	13	*rplS*	P02420	13	10.62		1
Stationary	13	*rplU*	P02422	11.56	9.85		2
Stationary	13	*rplV*	P02423	12.23	10.23		6
Stationary	13	*rplW*	P02424	11.2	9.94		3
Stationary	13	*rpmB*	P02428	8.88	11.42		4
Stationary	13	*rpsC*	P02352	25.85	10.27	100;63;54; 35;29	15
Stationary	13	*rpsD*	P02354	23.34	10.05	28;18	16
Stationary	13	*rpsE*	P02356	17.47	10.11	20	11
Stationary	13	*rpsF*	P02358	15.32	5.16		7
Stationary	13	*rpsG*	P02359	19.89	10.37		10
Stationary	13	*rpsI*	P02363	14.73	10.94		4
Stationary	13	*rpsM*	P02369	12.97	10.78		8
Stationary	13	*rpsR*	P02374	8.86	10.6		2
Stationary	13	*rpsS*	P02375	10.3	10.52		4
Stationary	13	*rpsT*	P02378	9.55	11.18		3
Stationary	14	*hupA*	P02342	9.53	9.57		4
Stationary	14	*hupB*	P02341	9.23	9.69		5
Stationary	14	*relA*	P11585	83.88	6.29		2
Stationary	14	*rmf*	P22986	6.51	10.86		1
Stationary	14	*rplB*	P02387	29.73	10.93		3
Stationary	14	*rplO*	P02413	14.98	11.18		5
Stationary	14	*rplP*	P02414	15.28	11.22		5
Stationary	14	*rplQ*	P02416	14.36	11.05	21	3
Stationary	14	*rplT*	P02421	13.37	11.47		2
Stationary	14	*rplU*	P02422	11.56	9.85		3
Stationary	14	*rplV*	P02423	12.23	10.23		4
Stationary	14	*rpmB*	P02428	8.88	11.42		2
Stationary	14	*rpsC*	P02352	25.85	10.27	28	10
Stationary	14	*rpsD*	P02354	23.34	10.05	28	11
Stationary	14	*rpsE*	P02356	17.47	10.11	18	5
Stationary	14	*rpsF*	P02358	15.32	5.16		7

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	14	*rpsG*	P02359	19.89	10.37	18	8
Stationary	14	*rpsI*	P02363	14.73	10.94		4
Stationary	14	*rpsK*	P02366	13.71	11.33		1
Stationary	14	*rpsM*	P02369	12.97	10.78		5
Stationary	14	*rpsR*	P02374	8.86	10.6		2
Stationary	14	*rpsT*	P02378	9.55	11.18		3
Stationary	14	*ybiB*	P30177	35.05	6.38		5
Stationary	15	*aidB*	P33224	60.59	7.98	54;47	7
Stationary	15	*hupA*	P02342	9.53	9.57		3
Stationary	15	*ihfA*	P06984	11.35	9.34		3
Stationary	15	*ihfB*	P08756	10.65	9.34		2
Stationary	15	*pcnB*	P13685	54.66	9.68	47	
Stationary	15	*rmf*	P22986	6.51	10.86		2
Stationary	15	*rplB*	P02387	29.73	10.93		8
Stationary	15	*rplD*	P02388	22.09	9.72		3
Stationary	15	*rplE*	P02389	20.17	9.49		4
Stationary	15	*rplM*	P02410	16.02	9.91		2
Stationary	15	*rplN*	P02411	13.54	10.43		4
Stationary	15	*rplO*	P02413	14.98	11.18		5
Stationary	15	*rplP*	P02414	15.28	11.22	18	5
Stationary	15	*rplQ*	P02416	14.36	11.05	18	5
Stationary	15	*rplS*	P02420	13	10.62		3
Stationary	15	*rplT*	P02421	13.37	11.47		3
Stationary	15	*rplU*	P02422	11.56	9.85		2
Stationary	15	*rplV*	P02423	12.23	10.23		7
Stationary	15	*rpmB*	P02428	8.88	11.42		2
Stationary	15	*rpsC*	P02352	25.85	10.27	29	11
Stationary	15	*rpsD*	P02354	23.34	10.05	28	10
Stationary	15	*rpsE*	P02356	17.47	10.11	21	5
Stationary	15	*rpsF*	P02358	15.32	5.16		7
Stationary	15	*rpsG*	P02359	19.89	10.37	23;21	9
Stationary	15	*rpsI*	P02363	14.73	10.94		5
Stationary	15	*rpsK*	P02366	13.71	11.33		4
Stationary	15	*rpsM*	P02369	12.97	10.78		7
Stationary	15	*rpsR*	P02374	8.86	10.6		2
Stationary	15	*rpsT*	P02378	9.55	11.18		5
Stationary	15	*ybiB*	P30177	35.05	6.38	36	11

Table A.1. Continued

Cell	Fraction#	Gene Name	SwissProt ID	MW (kD)	pI	In-Gel Obs.MW (kD)	LC Peptides#
Stationary	16	*aidB*	P33224	60.59	7.98	120;54	6
Stationary	16	*hisB*	P06987	40.28	5.76	42	3
Stationary	16	*hupA*	P02342	9.53	9.57		2
Stationary	16	*ihfA*	P06984	11.35	9.34		4
Stationary	16	*ihfB*	P08756	10.65	9.34		3
Stationary	16	*infC*	P02999	20.56	9.54		2
Stationary	16	*rplB*	P02387	29.73	10.93		4
Stationary	16	*rplD*	P02388	22.09	9.72		3
Stationary	16	*rplM*	P02410	16.02	9.91		1
Stationary	16	*rplN*	P02411	13.54	10.43		1
Stationary	16	*rplO*	P02413	14.98	11.18		5
Stationary	16	*rplP*	P02414	15.28	11.22		3
Stationary	16	*rplS*	P02420	13	10.62		2
Stationary	16	*rplT*	P02421	13.37	11.47		3
Stationary	16	*rplU*	P02422	11.56	9.85		2
Stationary	16	*rplV*	P02423	12.23	10.23		4
Stationary	16	*rpmB*	P02428	8.88	11.42		3
Stationary	16	*rpsC*	P02352	25.85	10.27	42;35;29;27	6
Stationary	16	*rpsD*	P02354	23.34	10.05	27	5
Stationary	16	*rpsE*	P02356	17.47	10.11	20	4
Stationary	16	*rpsF*	P02358	15.32	5.16		3
Stationary	16	*rpsG*	P02359	19.89	10.37	22;20	5
Stationary	16	*rpsI*	P02363	14.73	10.94		4
Stationary	16	*rpsK*	P02366	13.71	11.33		3
Stationary	16	*rpsM*	P02369	12.97	10.78		5
Stationary	16	*rpsR*	P02374	8.86	10.6		2
Stationary	16	*rpsT*	P02378	9.55	11.18		4
Stationary	16	*ybiB*	P30177	35.05	6.38		5

VITA

NAME:	Lili Niu
ADDRESS:	Department of Biochemistry & Biophysics 103 Biochemistry/Biophysics Building Texas A&M University 2128 TAMU College Station, Texas 77843-2128
BIRTHPLACE:	Beijing, P. R. China
EDUCATION:	Ph.D., Biochemistry (2007) Texas A&M University College Station, TX 77840 B.Sc., Biotechnology (2001) Beijing University Beijing, P. R. China